

A framework for the analytical performance assessment of matrix and tensor-based ESPRIT-type algorithms

Florian Roemer* *Student Member, IEEE* and Martin Haardt, *Senior Member, IEEE*

Abstract—In this paper we present a generic framework for the asymptotic performance analysis of subspace-based parameter estimation schemes. It is based on earlier results on an explicit first-order expansion of the estimation error in the signal subspace obtained via an SVD of the noisy observation matrix. We extend these results in a number of aspects. Firstly, we demonstrate that an explicit first-order expansion of the Higher-Order SVD (HOSVD)-based subspace estimate can be derived. Secondly, we show how to obtain explicit first-order expansions of the estimation error of arbitrary ESPRIT-type algorithms and provide the expressions for R -D matrix-based and tensor-based Standard ESPRIT as well as Unitary ESPRIT. Thirdly, we derive closed-form expressions for the mean square error (MSE) and show that they only depend on the second-order moments of the noise. Hence, we only need the noise to be zero mean and possess finite second order moments. Additional assumptions such as Gaussianity or circular symmetry are not needed. Fourthly, we investigate the effect of using Structured Least Squares (SLS) to solve the overdetermined shift invariance equations in ESPRIT and provide an explicit first-order expansion as well as a closed-form MSE expression. Finally, we simplify the MSE for the special case of a single source and compute the asymptotic efficiency of the investigated ESPRIT-type algorithms in compact closed-form expressions which only depend on the array size and the effective SNR.

Our results are more general than existing results on the performance analysis of ESPRIT-type algorithms since (a) we do not need any assumptions about the noise except for the mean to be zero and the second-order moments to be finite (in contrast to earlier results that require Gaussianity and/or second-order circular symmetry); (b) our results are asymptotic in the effective SNR, i.e., we do not require the number of samples to be large (in fact we can analyze even the single-snapshot case); (c) we present a framework that incorporates the SVD-based and the HOSVD-based subspace estimates as well as Structured Least Squares in one unified manner.

I. INTRODUCTION

HIGH resolution parameter estimation from R -dimensional (R -D) signals is a task required for a variety

Parts of this paper have been published at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011, at the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010)*, Dallas, TX, Mar. 2010, and at the *Third International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009)*, Aruba, Dutch Antilles, Dec. 2009.

The authors F. Roemer and M. Haardt are with Ilmenau University of Technology, Communications Research Laboratory, P. O. Box 10 05 65, D-98684 Ilmenau, Germany, e-mail: {florian.roemer, martin.haardt}@tu-ilmenau.de, phone: +49 (3677) 69-2613, fax: +49 (3677) 69-1195, WWW: <http://www.tu-ilmenau.de/crl>.

* corresponding author

of applications, such as estimating the multi-dimensional parameters of the dominant multipath components from MIMO channel measurements [12], which may be used for geometry-based channel modeling. Other applications include radar [24], wireless communications [20], sonar, seismology, and medical imaging. In [11], we have shown that in the R -D case ($R \geq 2$), tensors can be used to store and manipulate the R -D signals in their native multidimensional form. Based on this idea, we have proposed an enhanced tensor-based signal subspace estimate as well as ESPRIT-type algorithms based on tensors in [11]. Their superior performance was shown based on Monte-Carlo simulations.

In this paper we present a framework for the analytical performance assessment of subspace-based parameter estimation schemes and apply it to derive a first-order perturbation expansion for the tensor-based subspace estimate. Moreover, we find first-order expansions for the estimation error of ESPRIT-type algorithms and derive generic mean square error (MSE) expressions which only depend on the second-order moments of the noise and hence do not require Gaussianity or circular symmetry. This approach allows to assess the gain from using tensors instead of matrices analytically in order to determine in which scenarios it is particularly pronounced. We apply the framework for the analysis of R -D Standard ESPRIT, R -D Unitary ESPRIT, R -D Standard Tensor-ESPRIT and R -D Unitary Tensor-ESPRIT. Moreover, we investigate the effect of using Structured Least Squares (SLS) for the solution of the invariance equations. Finally, we present simplified MSE expressions for the special case of a single source impinging on a Uniform Linear Array (ULA) as well as a Uniform Rectangular Array (URA) and observed under circularly symmetric white noise. These expressions only depend on the effective Signal to Noise Ratio (SNR) and the array size and the allow to compute the asymptotic efficiency in closed-form.

Analytical performance assessment of subspace-based parameter estimation schemes has a long standing history in signal processing. Shortly after the publication of the most prominent candidates, MUSIC [32] and ESPRIT [31], analytical results on their performance have appeared. The most frequently cited papers are [14] for the MUSIC algorithm and [28] for ESPRIT. However, many follow-up papers exist which extend the original results, e.g., [26], [7], [23], [41], [22], and many others. However, these results have in common that they all go more or less directly back to a result on the distribution of the eigenvectors of a sample covariance matrix from [1], [2].

In contrast to these results, in [17] an entirely different approach was proposed, which provides an explicit first-order expansion of the subspace of a desired signal component if observed superimposed by a small additive perturbation. This approach has a number of advantages compared to [2]. Firstly, [2] is asymptotic in the sample size N , i.e., the result becomes only accurate as the number of snapshots N is very large, whereas [17] is asymptotic in the effective SNR, i.e., it can be used even for $N = 1$ as long as the noise variance is sufficiently small. Secondly, [2] requires strong Gaussianity assumptions, not only on the perturbation (i.e., the noise), but also on the source symbols. Since [17] is explicit, no assumptions about the statistics of either the desired signal or the perturbation are needed. Note that it has recently been shown that [2] can be extended to the non-Gaussian and the non-circular case in [5]. However, the large sample size assumption is still needed. Thirdly, the covariance expressions from [2] are much less intuitive than the expansion from [17] which shows directly how much of the noise subspace “leaks into” the signal subspace due to the erroneous estimate. Finally, the expressions involved in [2] are quite complex and tough to handle, whereas [17] requires only a few terms which appear directly as block matrices of the SVD of the noise-free observation matrix.

Due to these advantages we clearly favor [17] as a starting point. The authors in [17] have already shown that their results on the perturbation of the subspace can be used to find a first order expansion for the MUSIC, the Root-MUSIC, the Min-Norm, the State-Space-Realization, and even the ESPRIT algorithm. However, they only considered 1-D Standard ESPRIT. We extend their work by considering multiple dimensions (R -D ESPRIT), by incorporating forward-backward-averaging (for Unitary ESPRIT), by considering the tensor-based subspace estimate (for Standard and Unitary Tensor-ESPRIT), by investigating the effect of using Structured Least Squares (SLS) to solve the invariance equation instead of the Least Squares (LS) solution used in [17], and by providing generic mean square error (MSE) expressions of the resulting estimation errors in these cases. Note that our MSE expressions depend on the second-order moments of the noise only. Hence we only assume it to be zero mean (due to the asymptotic nature of our performance analysis), but do not require it to be Gaussian distributed, white, or circularly symmetric. This is a particularly attractive feature of our approach with respect to different types of preprocessing which alters the noise statistics, e.g., spatial smoothing (which yields spatially correlated noise) or forward-backward averaging (which annihilates the circular symmetry of the noise). Since we do not require spatial whiteness or circular symmetry, our MSE expressions are directly applicable to a wide range of ESPRIT-type algorithms.

There have been other follow-up papers based on [17]. For instance, [40] provides a first-order and second-order perturbation expansion which can be seen as a generalization of [17]. In [19] the authors show that there is also a first-order contribution of the perturbation of the signal subspace which lies in the signal subspace (which [40] and [17] have argued to be of second order and hence negligible). Note that a

perturbation expansions for the signal subspace and null space projectors based on the sample covariance matrix is provided in [15] where the expansion up to an arbitrary order is derived based on a recurrence relation. A mean square error expression for Standard ESPRIT is provided in [18], however, it does not generalize easily to the tensor case and it assumes circular symmetry of the noise. Note that the latter assumption implies that it is not applicable to Unitary ESPRIT, since Forward-Backward Averaging annihilates the circular symmetry of the noise. Moreover, other authors have studied the asymptotical performance of ESPRIT, e.g., [35], [6] where harmonic retrieval from time series is investigated and MSE expressions for a large number of snapshots as well as MSE expressions for a high SNR are derived. Note that we find MSE expressions compatible to [6] by only assuming a high effective SNR, i.e., either the number of snapshots or the SNR can tend to infinity. Interestingly, [35], [6] also consider the special case for a single source. However, the expressions provided there are specific to harmonic retrieval from time series and they are not compared to the corresponding Cramér-Rao Bound. Some analytical results on the asymptotic efficiency of MUSIC, Root-MUSIC, ESPRIT, and TLS-ESPRIT are, among others, presented in [27], [28], [29], and [25], respectively. However, these results are asymptotic in the number of snapshots N and sometimes even in the number of sensors M . The asymptotic equivalence of LS-ESPRIT, TLS-ESPRIT, Pro-ESPRIT, and the Matrix Pencil method has also been shown, see for instance [13]. Overall, in the matrix case, the number of existing results is quite large, since the underlying methods have been known for more than two decades. However, concerning the tensor case and the incorporation of Structured Least Squares, the existing results are much more scarce.

In the tensor case a first-order expansion for the HOSVD has been proposed in [4]. However, it is not suitable for our application since it does not consider the HOSVD-based subspace estimate but the subspaces of the separate n -mode unfoldings and their singular values. Moreover, the perturbation is modeled via a single scalar real-valued parameter ϵ . A first-order expansion for the best rank- (R_1, R_2, R_3) -expansion is provided in [3]. However, again, it is not directly applicable for analyzing the HOSVD-based subspace estimate as it investigates the approximation error of the entire tensor. Consequently, our approach to analyze the HOSVD-based subspace estimate based on the link to the SVD-based subspace estimate via a structured projection is entirely novel. Moreover, the application of these results to find the analytical performance of Tensor-ESPRIT-type algorithms is novel as well. It is a particular strength of the framework we use that many extensions and modifications of ESPRIT are easily incorporated, e.g., Forward-Backward-Averaging or Structured Least Squares.

This paper is organized as follows: The notation and the data model are introduced in Sections II and III, respectively. The subsequent Section IV reviews the first-order perturbation of the matrix-based subspace estimate and presents the extension to the tensor case. The performance analysis of ESPRIT-type algorithms is shown in V. Numerical results are presented in Section VI before drawing the conclusions in Section VII.

II. NOTATION

In order to facilitate the distinction between scalars, matrices, and tensors, the following notation is used: Scalars are denoted as italic letters ($a, b, \dots, A, B, \dots, \alpha, \beta, \dots$), column vectors as lower-case bold-face letters ($\mathbf{a}, \mathbf{b}, \dots$), matrices as bold-face capitals ($\mathbf{A}, \mathbf{B}, \dots$), and tensors are written as bold-face calligraphic letters ($\mathcal{A}, \mathcal{B}, \dots$). Lower-order parts are consistently named: the (i, j) -element of the matrix \mathbf{A} , is denoted as $a_{i,j}$ and the (i, j, k) -element of a third order tensor \mathcal{B} as $b_{i,j,k}$.

We use the superscripts $\text{T}, \text{H}, *, ^{-1}, +$ for transposition, Hermitian transposition, complex conjugation, matrix inversion, and the Moore-Penrose pseudo inverse of a matrix, respectively. The trace of a matrix \mathbf{A} is written as $\text{Tr}(\mathbf{A})$. Moreover, the Kronecker product of two matrices \mathbf{A} and \mathbf{B} is denoted as $\mathbf{A} \otimes \mathbf{B}$ and the Khatri-Rao product (column-wise Kronecker product) as $\mathbf{A} \diamond \mathbf{B}$. The operator $\text{vec}\{\mathbf{A}\}$ stacks the column of a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ into a column vector of length $M \cdot N \times 1$. It satisfies the following property

$$\text{vec}\{\mathbf{A} \cdot \mathbf{X} \cdot \mathbf{B}\} = (\mathbf{B}^{\text{T}} \otimes \mathbf{A}) \cdot \text{vec}\{\mathbf{X}\}. \quad (1)$$

An n -mode vector of an $(I_1 \times I_2 \times \dots \times I_N)$ -dimensional tensor \mathcal{A} is an I_n -dimensional vector obtained from \mathcal{A} by varying the index i_n and keeping the other indices fixed. Moreover, a matrix unfolding of the tensor \mathcal{A} along the n -th mode is denoted by $[\mathcal{A}]_{(n)}$ and can be understood as a matrix containing all the n -mode vectors of the tensor \mathcal{A} . The order of the columns is chosen in accordance with [4].

The outer product of the tensors $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \dots \times J_M}$ is given by

$$\mathcal{C} = \mathcal{A} \circ \mathcal{B} \in \mathbb{C}^{I_1 \times \dots \times I_N \times J_1 \times \dots \times J_M}, \quad \text{where} \quad (2)$$

$$c_{i_1, i_2, \dots, i_N, j_1, j_2, \dots, j_M} = a_{i_1, i_2, \dots, i_N} \cdot b_{j_1, j_2, \dots, j_M}.$$

In other words, the tensor \mathcal{C} contains all possible combinations of pairwise products between the elements of \mathcal{A} and \mathcal{B} . This operator is very closely related to the Kronecker product defined for matrices.

The n -mode product of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and a matrix $\mathbf{U} \in \mathbb{C}^{J_n \times I_n}$ along the n -th mode is denoted as $\mathcal{B} = \mathcal{A} \times_n \mathbf{U}$ and defined via

$$\mathcal{B} = \mathcal{A} \times_n \mathbf{U} \Leftrightarrow [\mathcal{B}]_{(n)} = \mathbf{U} \cdot [\mathcal{A}]_{(n)}, \quad (3)$$

i.e., it may be visualized by multiplying all n -mode vectors of \mathcal{A} from the left-hand side by the matrix \mathbf{U} . Note that the n -mode product satisfies

$$(\mathcal{A} \times_r \mathbf{U}_r) \times_r \mathbf{V}_r = \mathcal{A} \times_r (\mathbf{V}_r \cdot \mathbf{U}_r) \quad (4)$$

$$[\mathcal{A} \times_1 \mathbf{U}_1 \dots \times_R \mathbf{U}_R]_{(r)} = \mathbf{U}_r \cdot [\mathcal{A}]_{(r)}. \quad (5)$$

$$(\mathbf{U}_{r+1} \otimes \dots \otimes \mathbf{U}_R \otimes \mathbf{U}_1 \otimes \dots \otimes \mathbf{U}_{r-1})^{\text{T}} \quad (5)$$

for $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, $\mathbf{U}_r \in \mathbb{C}^{J_r \times I_r}$ and $\mathbf{V}_r \in \mathbb{C}^{K_r \times J_r}$.

The higher-order SVD (HOSVD) [4] of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is given by

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \dots \times_N \mathbf{U}_N, \quad (6)$$

where $\mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is the core tensor which satisfies the all-orthogonality conditions [4] and $\mathbf{U}_n \in \mathbb{C}^{I_n \times I_n}$, $n =$

$1, 2, \dots, N$, are the unitary matrices of n -mode singular vectors.

We also define the concatenation of two tensors along the n -th mode via the operator $[\mathcal{A} \sqcup_n \mathcal{B}]$. The Euclidean (vector) norm, the Frobenius (matrix) norm, and the Higher-Order Frobenius (tensor) norm are denoted by $\|\mathbf{a}\|_2$, $\|\mathbf{A}\|_{\text{F}}$, and $\|\mathcal{A}\|_{\text{H}}$, respectively. All three norms are computed by taking the square-root of the sum of the squared magnitude of all the elements in their arguments.

The matrix $\mathbf{K}_{M \times N}$ denotes the commutation matrices [21] which satisfy

$$\mathbf{K}_{M \times N} \cdot \text{vec}\{\mathbf{A}^{\text{T}}\} = \text{vec}\{\mathbf{A}\} \quad (7)$$

$$\mathbf{K}_{M \times P}^{\text{T}} \cdot (\mathbf{A} \otimes \mathbf{B}) \cdot \mathbf{K}_{N \times Q} = \mathbf{B} \otimes \mathbf{A} \quad (8)$$

for $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\mathbf{B} \in \mathbb{C}^{P \times Q}$.

A projection matrix onto the column space of a matrix $\mathbf{A} \in \mathbb{C}^{M \times r}$ is denoted as $\text{Pr}_{\mathbf{A}} = \mathbf{A} \cdot \mathbf{A}^+ \in \mathbb{C}^{M \times M}$ and its orthogonal complement by $\text{Pr}_{\mathbf{A}}^{\perp} = \mathbf{I}_M - \text{Pr}_{\mathbf{A}}$. Note that for $r = 1$, i.e., $\mathbf{A} = \mathbf{a}$, this matrix can be computed as $\text{Pr}_{\mathbf{a}} = \frac{\mathbf{a} \cdot \mathbf{a}^{\text{H}}}{\mathbf{a}^{\text{H}} \cdot \mathbf{a}}$.

A $p \times p$ matrix \mathbf{Q}_p is called left- Π -real if $\Pi_p \cdot \mathbf{Q}_p^* = \mathbf{Q}_p$, where Π_p is the $p \times p$ exchange matrix with ones on its antidiagonal and zeros elsewhere. The special set of unitary sparse left- Π -real matrices introduced in [9] is denoted as $\mathbf{Q}_p^{(\text{s})}$. Furthermore, a matrix $\mathbf{X} \in \mathbb{C}^{M \times N}$ is called centro-Hermitian if $\Pi_M \cdot \mathbf{X}^* \cdot \Pi_N = \mathbf{X}$. The vector \mathbf{e}_k denotes the k -th column of an identity matrix.

III. DATA MODEL

A. Matrix-based and tensor-based data model

The observations are modeled as a superposition of d undamped exponentials sampled on an R -dimensional grid of size $M_1 \times M_2 \times \dots \times M_R$ at N subsequent time instants [10]. The measurement samples are given by

$$x_{m_1, m_2, \dots, m_R, t_n} = \sum_{i=1}^d s_i(t_n) \prod_{r=1}^R e^{j \cdot (m_r - 1) \cdot \mu_i^{(r)} + n_{m_1, m_2, \dots, m_R, t_n}}, \quad (9)$$

where $m_r = 1, 2, \dots, M_r$, $n = 1, 2, \dots, N$, $s_i(t_n)$ denotes the complex amplitude of the i -th exponential at time instant t_n , $\mu_i^{(r)}$ symbolizes the spatial frequency of the i -th exponential in the r -th mode for $i = 1, 2, \dots, d$ and $r = 1, 2, \dots, R$, and $n_{m_1, m_2, \dots, m_R, t_n}$ represents the zero mean additive noise component inherent in the measurement process¹. In the context of array signal processing, each of the R -dimensional exponentials represents one planar wavefront and the complex amplitudes $s_i(t_n)$ are the symbols. It is our goal to estimate the spatial frequencies $\mu_i^{(r)}$ for $r = 1, 2, \dots, R$, $i = 1, 2, \dots, d$ and their correct pairing.

In order to arrive at a more compressed formulation of the data model in (9) we collect the samples $x_{m_1, m_2, \dots, m_R, t_n}$ into one array. As our signal is referenced by $R + 1$ indices, the

¹Note that equation (9) assumes a uniform sampling in the spatial domain. However, this assumption can be relaxed to more generic geometries as long as they feature shift invariances and can be constructed as the outer product of R one-dimensional sampling grids.

most natural way of formulating the model is to employ an $(R+1)$ -way array $\mathcal{X} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$ which contains $x_{m_1, m_2, \dots, m_R, t_n}$ for $m_r = 1, 2, \dots, M_r$, $r = 1, 2, \dots, R$, and $n = 1, 2, \dots, N$. We can then conveniently express \mathcal{X} as [11]

$$\mathcal{X} = \mathcal{A} \times_{R+1} \mathbf{S}^T + \mathcal{N}, \quad (10)$$

where $\mathbf{S} \in \mathbb{C}^{d \times N}$ contains the amplitudes $s_i[n]$ and $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times N}$ collects all the noise samples $n_{m_1, m_2, \dots, m_R, t_n}$ in the same manner as \mathcal{X} . Finally, $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$ is referred to as the “array steering tensor” [11]. It can be expressed by virtue of the concatenation operator via

$$\mathcal{A} = [\mathcal{A}_1 \sqcup_{R+1} \mathcal{A}_2 \sqcup_{R+1} \dots \sqcup_{R+1} \mathcal{A}_d] \quad (11)$$

$$\mathcal{A}_i = \mathbf{a}^{(1)}(\mu_i^{(1)}) \circ \mathbf{a}^{(2)}(\mu_i^{(2)}) \circ \dots \circ \mathbf{a}^{(R)}(\mu_i^{(R)}) \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_R \times d}$$

where $\mathbf{a}^{(r)}(\mu_i^{(r)}) \in \mathbb{C}^{M_r \times 1}$ represents the array steering vector of the i -th source in the r -th mode.

An alternative expression for the array steering tensor is given by

$$\mathcal{A} = \mathcal{I}_{R+1, d} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \dots \times_R \mathbf{A}^{(R)}, \quad (12)$$

where $\mathbf{A}^{(r)} = [\mathbf{a}^{(r)}(\mu_1^{(r)}) \dots \mathbf{a}^{(r)}(\mu_d^{(r)})] \in \mathbb{C}^{M_r \times d}$ is referred to as the array steering matrix in the r -th mode.

The strength of the data model in (10) is that it represents the signal in its natural multidimensional structure by virtue of the measurement tensor \mathcal{X} . Before tensor calculus was used in this area, a matrix-based formulation of (10) was needed. This requires stacking some of the dimensions into rows or columns. A meaningful definition of a measurement matrix \mathbf{X} is to apply stacking to all “spatial” dimensions $1, 2, \dots, R$ along the rows and align the snapshots $n = 1, 2, \dots, N$ as the columns. Mathematically, we can write $\mathbf{X} = [\mathcal{X}]_{(R+1)}^T \in \mathbb{C}^{M \times N}$, where $M = \prod_{r=1}^R M_r$. Applying this stacking operation to (10), we arrive at the matrix-based data model [10]

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}. \quad (13)$$

Here, $\mathbf{A} = [\mathcal{A}]_{(R+1)}^T \in \mathbb{C}^{M \times d}$ and $\mathbf{N} = [\mathcal{N}]_{(R+1)}^T \in \mathbb{C}^{M \times N}$. Note that \mathbf{A} is highly structured since it satisfies

$$\mathbf{A} = [\mathcal{A}]_{(R+1)}^T = \mathbf{A}^{(1)} \diamond \mathbf{A}^{(2)} \diamond \dots \diamond \mathbf{A}^{(R)}. \quad (14)$$

B. Subspace estimation

The first step in all subspace-based parameter estimation schemes is the estimation of a basis for the signal subspace from the noisy observations. In the matrix case, this can, for instance, be achieved by a truncated SVD of \mathbf{X} . Let $\hat{\mathbf{U}}_s \in \mathbb{C}^{M \times d}$ be the matrix containing the d dominant left singular vectors of \mathbf{X} . Then the column space of $\hat{\mathbf{U}}_s$ is an estimate for the signal subspace spanned by the columns of \mathbf{A} and we can write $\mathbf{A} \approx \hat{\mathbf{U}}_s \cdot \mathbf{T}$ for a non-singular matrix $\mathbf{T} \in \mathbb{C}^{d \times d}$.

A tensor-based extension of this subspace estimate was proposed in [11]. To this end, let the truncated HOSVD of \mathcal{X} be given by

$$\mathcal{X} \approx \hat{\mathcal{S}} \times_1 \hat{\mathbf{U}}_1^{[s]} \dots \times_R \hat{\mathbf{U}}_R^{[s]} \times_{R+1} \hat{\mathbf{U}}_{R+1}^{[s]}, \quad (15)$$

where $\hat{\mathcal{S}}^{[s]} \in \mathbb{C}^{p_1 \times \dots \times p_R \times d}$ is the truncated core tensor and $\hat{\mathbf{U}}_r^{[s]} \in \mathbb{C}^{M_r \times p_r}$ for $r = 1, 2, \dots, R$, $\hat{\mathbf{U}}_{R+1}^{[s]} \in \mathbb{C}^{N \times d}$ are the matrices of dominant r -mode singular vectors. Here, $p_r = \text{rank}([\mathcal{X}_0]_{(r)})$ represents the r -rank of the noise-free observation tensor $\mathcal{X}_0 = \mathcal{A} \times_{R+1} \mathbf{S}^T$. Based on (15), a tensor-based subspace estimate can be defined as

$$\hat{\mathcal{U}}^{[s]} = \hat{\mathcal{S}}^{[s]} \times_1 \hat{\mathbf{U}}_1^{[s]} \dots \times_R \hat{\mathbf{U}}_R^{[s]} \times_{R+1} \hat{\mathbf{\Sigma}}_s^{-1}. \quad (16)$$

Note that the $(R+1)$ -mode multiplication with $\hat{\mathbf{\Sigma}}_s^{-1}$ is introduced in addition to its original definition in [11] since it simplifies the notation we need at this point and it has no impact on the subspace estimation accuracy. Here $\hat{\mathbf{\Sigma}}_s$ refers to the diagonal matrix containing the d dominant singular values of \mathbf{X} on its main diagonal.

Note that $\hat{\mathcal{U}}^{[s]}$ satisfies $\hat{\mathcal{U}}^{[s]} \approx \mathcal{A} \times_{R+1} \bar{\mathbf{T}}$ for a non-singular matrix $\bar{\mathbf{T}} \in \mathbb{C}^{d \times d}$. Based on $\hat{\mathcal{U}}^{[s]}$, an improved signal subspace estimate is given by the matrix $[\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T \in \mathbb{C}^{M \times d}$.

IV. PERTURBATIONS OF THE SUBSPACE ESTIMATES

A. Review of perturbation results for the SVD

Let us first review the results from [17] which are relevant to the discussion in this section. Let $\mathbf{X}_0 = \mathbf{A} \cdot \mathbf{S} \in \mathbb{C}^{M \times N}$ be a matrix containing the noise-free observations such that $\mathbf{X} = \mathbf{X}_0 + \mathbf{N}$ where \mathbf{N} represents the undesired perturbation (noise).

The SVD of \mathbf{X}_0 can be expressed as

$$\mathbf{X}_0 = [\mathbf{U}_s \quad \mathbf{U}_n] \cdot \begin{bmatrix} \mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot [\mathbf{V}_s \quad \mathbf{V}_n]^H, \quad (17)$$

where the columns of $\mathbf{U}_s \in \mathbb{C}^{M \times d}$ provide an orthonormal basis for the signal subspace which we want to estimate. Moreover $\mathbf{\Sigma}_s = \text{diag}([\sigma_1, \sigma_2, \dots, \sigma_d]) \in \mathbb{R}^{d \times d}$ contains the d non-zero singular values on its main diagonal. We find an estimate for \mathbf{U}_s by computing an SVD of the noisy observation matrix \mathbf{X} which can be expressed as

$$\mathbf{X} = [\hat{\mathbf{U}}_s \quad \hat{\mathbf{U}}_n] \cdot \begin{bmatrix} \hat{\mathbf{\Sigma}}_s & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Sigma}}_n \end{bmatrix} \cdot [\hat{\mathbf{V}}_s \quad \hat{\mathbf{V}}_n]^H, \quad (18)$$

where the “hat” denotes the estimated quantities. We can write $\hat{\mathbf{U}}_s = \mathbf{U}_s + \Delta \mathbf{U}_s$, where $\Delta \mathbf{U}_s$ represents the estimation error. At this point we are ready to state the main result on the first order perturbation expansion of $\Delta \mathbf{U}_s$ from [17]

$$\Delta \mathbf{U}_s = \mathbf{U}_n \cdot \mathbf{\Gamma}_n + \mathcal{O}\{\Delta^2\}, \text{ where } \Delta = \|\mathbf{N}\| \text{ and } \mathbf{\Gamma}_n = \mathbf{U}_n^H \cdot \mathbf{N} \cdot \mathbf{V}_s \cdot \mathbf{\Sigma}_s^{-1} \in \mathbb{C}^{(M-d) \times d} \quad (19)$$

Here $\|\cdot\|$ represents an arbitrary sub-multiplicative² norm, e.g., the Frobenius norm. Equation (19) shows the first order expansion of the signal subspace estimation error $\Delta \mathbf{U}_s$ in terms of the noise subspace \mathbf{U}_n , i.e., how much of the noise subspace “leaks into” the signal subspace due to the estimation errors from the perturbation \mathbf{N} . Since it is explicit in \mathbf{N} it

²A matrix norm is called submultiplicative if $\|\mathbf{A} \cdot \mathbf{B}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$ for arbitrary matrices \mathbf{A} and \mathbf{B} .

makes no assumptions about the *statistics* of N , in fact, it is purely deterministic.

The expansion (19) only models the leakage of the noise subspace into the signal subspace. That is to say, the perturbation of the particular basis (the columns of \mathbf{U}_s) is ignored. While for subspace-based parameter estimation schemes this is indeed sufficient since the particular choice of the basis is irrelevant, there are other applications where this term matters. For instance, in a communication system where the channel is decomposed into its individual eigenmodes and one or several of these eigenmodes are used for transmission, such errors have a major impact. Therefore, other authors have extended (19) to take this error term into account. For instance, in [19] the authors provide the following expansion

$$\Delta \mathbf{U}_s = \mathbf{U}_n \cdot \mathbf{\Gamma}_n + \mathbf{U}_s \cdot \mathbf{\Gamma}_s + \mathcal{O}\{\Delta^2\}, \text{ where} \quad (20)$$

$$\mathbf{\Gamma}_s = \mathbf{D} \odot \left(\mathbf{U}_s^H \cdot \mathbf{N} \cdot \mathbf{V}_s \cdot \mathbf{\Sigma}_s + \mathbf{\Sigma}_s \cdot \mathbf{V}_s^H \cdot \mathbf{N}^H \cdot \mathbf{U}_s \right) \in \mathbb{C}^{p_r \times p_r}.$$

Here, the matrix \mathbf{D} is defined as

$$[\mathbf{D}]_{(k,\ell)} = \begin{cases} \frac{1}{\sigma_\ell^2 - \sigma_k^2} & k \neq \ell \\ 0 & k = \ell \end{cases} \text{ for } k, \ell = 1, 2, \dots, d. \quad (21)$$

Equation (20) additionally shows the perturbation of the individual singular vectors via the term $\mathbf{U}_s \cdot \mathbf{\Gamma}_s$. This term can be dropped for the evaluation of subspace-based parameter estimation schemes since for these, the particular choice of the basis is irrelevant. Therefore we do not consider it in Section V where ESPRIT-type algorithms are investigated. However, we show its impact in the simulation results in Section VI where the subspace estimation accuracy is evaluated.

B. Extension to the HOSVD-based subspace estimate

As we have shown in Section III-B, in the multidimensional case, an improved signal subspace estimate can be computed via the HOSVD of the measurement tensor \mathcal{X} . Since the HOSVD is computed via SVDs of the unfoldings, we can apply the same framework to find a perturbation expansion of the HOSVD-based subspace estimate. In order to distinguish unperturbed from estimated (perturbed) quantities we express \mathcal{X} as $\mathcal{X} = \mathcal{X}_0 + \mathcal{N}$, where $\mathcal{X}_0 = \mathcal{A} \times_{R+1} \mathbf{S}^T$ is the unperturbed observation tensor. The SVD of the r -mode unfoldings of \mathcal{X} and \mathcal{X}_0 are then given by

$$[\mathcal{X}_0]_{(r)} = [\mathbf{U}_r^{[s]} \quad \mathbf{U}_r^{[n]}] \cdot \begin{bmatrix} \mathbf{\Sigma}_r^{[s]} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot [\mathbf{V}_r^{[s]} \quad \mathbf{V}_r^{[n]}]^H \quad (22)$$

$$[\mathcal{X}]_{(r)} = [\hat{\mathbf{U}}_r^{[s]} \quad \hat{\mathbf{U}}_r^{[n]}] \cdot \begin{bmatrix} \hat{\mathbf{\Sigma}}_r^{[s]} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Sigma}}_r^{[n]} \end{bmatrix} \cdot [\hat{\mathbf{V}}_r^{[s]} \quad \hat{\mathbf{V}}_r^{[n]}]^H \quad (23)$$

where $\mathbf{\Sigma}_r^{[s]} = \text{diag}([\sigma_1^{(r)}, \sigma_2^{(r)}, \dots, \sigma_d^{(r)}])$ and $r = 1, 2, \dots, R$. Note that since (22) and (23) are in fact SVDs, we can apply (20) and find $\hat{\mathbf{U}}_r^{[s]} = \mathbf{U}_r^{[s]} + \Delta \mathbf{U}_r^{[s]}$ where

$$\Delta \mathbf{U}_r^{[s]} = \mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} + \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]} + \mathcal{O}\{\Delta^2\}, \quad (24)$$

$$\mathbf{\Gamma}_r^{[n]} = \mathbf{U}_r^{[n]H} \cdot [\mathcal{N}]_{(r)} \cdot \mathbf{V}_r^{[s]} \cdot \mathbf{\Sigma}_r^{[s]-1},$$

$$\mathbf{\Gamma}_r^{[s]} = \mathbf{D}_r \odot \left(\mathbf{U}_r^{[s]H} \cdot [\mathcal{N}]_{(r)} \cdot \mathbf{V}_r^{[s]} \cdot \mathbf{\Sigma}_r^{[s]} \right)$$

$$+ \mathbf{\Sigma}_r^{[s]} \cdot \mathbf{V}_r^{[s]H} \cdot [\mathcal{N}]_{(r)}^H \cdot \mathbf{U}_r^{[s]})$$

$$[\mathbf{D}_r]_{(k,\ell)} = \begin{cases} \frac{1}{\sigma_\ell^{(r)2} - \sigma_k^{(r)2}} & k \neq \ell \\ 0 & k = \ell \end{cases} \text{ for } k, \ell = 1, 2, \dots, d.$$

Our goal is to use the perturbation of the r -mode unfoldings to find a corresponding expansion for the HOSVD-based subspace estimate introduced in Section III-B. This is facilitated by the following theorem:

Theorem 1. *The HOSVD-based subspace estimate $[\hat{\mathbf{u}}^{[s]}]_{(R+1)}^T$ defined in (16) is linked to the SVD-based subspace estimate $\hat{\mathbf{U}}_s$ via the following algebraic relation*

$$[\hat{\mathbf{u}}^{[s]}]_{(R+1)}^T = (\hat{\mathbf{T}}_1 \otimes \hat{\mathbf{T}}_2 \otimes \dots \otimes \hat{\mathbf{T}}_R) \cdot \hat{\mathbf{U}}_s, \quad (25)$$

where $\hat{\mathbf{T}}_r \in \mathbb{C}^{M_r \times M_r}$ represent estimates of the projection matrices onto the r -spaces of \mathcal{X}_0 , which are computed via $\hat{\mathbf{T}}_r = \hat{\mathbf{U}}_r^{[s]} \hat{\mathbf{U}}_r^{[s]H}$.

Proof: Relation (25) was shown in [30] for $R = 2$. The proof for $R > 2$ proceeds in an analogous manner and is presented in Appendix A.

Equation (25) shows that a perturbation expansion for $[\hat{\mathbf{u}}^{[s]}]_{(R+1)}^T$ can be developed based on the subspaces of all $R + 1$ unfoldings, as the core tensor is not needed for its computation. The result is shown in the following theorem:

Theorem 2. *The HOSVD-based signal subspace estimate can be written as $[\hat{\mathbf{u}}^{[s]}]_{(R+1)}^T = \mathbf{U}_s + [\Delta \hat{\mathbf{u}}^{[s]}]_{(R+1)}^T$, where*

$$[\Delta \hat{\mathbf{u}}^{[s]}]_{(R+1)}^T = (\mathbf{T}_1 \otimes \mathbf{T}_2 \otimes \dots \otimes \mathbf{T}_R) \cdot \Delta \mathbf{U}_s$$

$$+ \left([\Delta \mathbf{U}_1^{[s]} \cdot \mathbf{U}_1^{[s]H}] \otimes \mathbf{T}_2 \otimes \dots \otimes \mathbf{T}_R \right) \cdot \mathbf{U}_s$$

$$+ \left(\mathbf{T}_1 \otimes [\Delta \mathbf{U}_2^{[s]} \cdot \mathbf{U}_2^{[s]H}] \otimes \dots \otimes \mathbf{T}_R \right) \cdot \mathbf{U}_s$$

$$+ \dots$$

$$+ \left(\mathbf{T}_1 \otimes \mathbf{T}_2 \otimes \dots \otimes [\Delta \mathbf{U}_R^{[s]} \cdot \mathbf{U}_R^{[s]H}] \right) \cdot \mathbf{U}_s$$

$$+ \mathcal{O}\{\Delta^2\}, \quad (26)$$

the SVD-based signal subspace perturbation $\Delta \mathbf{U}_s$ is given by (20) and the perturbation of the r -space can be computed via

$$\Delta \mathbf{U}_r^{[s]} = \mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} = \mathbf{U}_r^{[n]} \cdot \mathbf{U}_r^{[n]H} \cdot [\mathcal{N}]_{(r)} \cdot \mathbf{V}_r^{[s]} \cdot \mathbf{\Sigma}_r^{[s]-1}. \quad (27)$$

Proof: cf. Appendix B.

Note that while $\Delta \mathbf{U}_s$ in general contains both perturbation terms $\mathbf{U}_n \cdot \mathbf{\Gamma}_n$ and $\mathbf{U}_s \cdot \mathbf{\Gamma}_s$, for $\Delta \mathbf{U}_r^{[s]}$ the term $\mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]}$ cancels. This is not surprising since the r -mode subspaces enter (25) only via projection matrices for which the choice of the particular basis is irrelevant.

V. ASYMPTOTICAL ANALYSIS OF THE PARAMETER ESTIMATION ACCURACY

A. Review of perturbation results for the 1-D Standard ESPRIT

In [17] the authors point out that once a first order expansion of the subspace estimation error is available it can be used to find a corresponding first order expansion of the estimation error of a suitable parameter estimation scheme. One of the examples the authors show is the 1-D Standard ESPRIT algorithm using LS, which we use as a starting point to discuss various ESPRIT-type algorithms in this section. In the noise-free case, the shift invariance equation for 1-D Standard ESPRIT can be expressed as

$$\mathbf{J}_1 \cdot \mathbf{U}_s \cdot \boldsymbol{\Psi} = \mathbf{J}_2 \cdot \mathbf{U}_s \quad (28)$$

where $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{R}^{M^{(\text{sel})} \times M}$ are the selection matrices that select the $M^{(\text{sel})}$ elements from the M antenna elements which correspond to the first and the second subarray, respectively. Moreover, $\boldsymbol{\Psi} = \mathbf{Q} \cdot \boldsymbol{\Phi} \cdot \mathbf{Q}^{-1}$, where $\boldsymbol{\Phi} = \text{diag}([e^{j\mu_1}, \dots, e^{j\mu_d}]) \in \mathbb{C}^{d \times d}$ contains the spatial frequencies μ_k , $k = 1, 2, \dots, d$ that we want to estimate. Therefore, $\mu_k = \arg(\text{EV}_k\{\boldsymbol{\Psi}\})$, i.e., the k -th spatial frequency is obtained from the phase of the k -th eigenvalue ($\text{EV}_k\{\cdot\}$) of $\boldsymbol{\Psi}$.

In presence of noise, we only have an estimate $\hat{\mathbf{U}}_s$ of the signal subspace \mathbf{U}_s . Consequently, (28) does in general not have an exact solution anymore. A simple way of finding an approximate $\hat{\boldsymbol{\Psi}}$ is given by the LS solution which can be expressed as

$$\hat{\boldsymbol{\Psi}}_{\text{LS}} = (\mathbf{J}_1 \cdot \hat{\mathbf{U}}_s)^+ \cdot \mathbf{J}_2 \cdot \hat{\mathbf{U}}_s \quad (29)$$

To simplify the notation we skip the index “LS” for the remainder of this section (since only LS is considered) and pick it up again in the next section where we expand the discussion to SLS.

For the estimation error of the k -th spatial frequency corresponding to the LS solution from (29), [17] provides the following expansion

$$\Delta\mu_k = \text{Im}\left\{ \mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot [\mathbf{J}_2 / \lambda_k - \mathbf{J}_1] \cdot \Delta\mathbf{U}_s \cdot \mathbf{q}_k \right\} + \mathcal{O}\{\Delta^2\} \quad (30)$$

where $\lambda_k = e^{j\mu_k}$ and \mathbf{q}_k is the k -th column of \mathbf{Q} . Moreover, \mathbf{p}_k^T represents the k -th row vector of the matrix $\mathbf{P} = \mathbf{Q}^{-1}$. Note that $\Delta\mathbf{U}_s$ can be expanded in terms of the perturbation term \mathbf{N} by directly using the expansion (19). The additional term from (20) is not needed as it is irrelevant for the performance of ESPRIT.

B. Extension to R -D Standard (Tensor-)ESPRIT

The previous result from [17] on the first order perturbation expansion of 1-D Standard ESPRIT using LS is easily generalized to the R -D case. The reason is that for R -D LS-based ESPRIT, the R shift invariance equations are

solved independently from each other³. Hence, the arguments from [17] are readily applied to all modes individually and we directly obtain a first order expansion for the estimation error of the k -th spatial frequency in the r -th mode

$$\Delta\mu_k^{(r)} = \text{Im}\left\{ \mathbf{p}_k^T \cdot (\tilde{\mathbf{J}}_1^{(r)} \cdot \mathbf{U}_s)^+ \cdot [\tilde{\mathbf{J}}_2^{(r)} / \lambda_k^{(r)} - \tilde{\mathbf{J}}_1^{(r)}] \cdot \Delta\mathbf{U}_s \cdot \mathbf{q}_k \right\} + \mathcal{O}\{\Delta^2\} \quad (31)$$

where $\tilde{\mathbf{J}}_1^{(r)}, \tilde{\mathbf{J}}_2^{(r)} \in \mathbb{R}^{M_r^{(\text{sel})} \times M}$ are the effective R -D selection matrix for the first and the second subarray in the r -th mode, respectively. They can be expressed as $\tilde{\mathbf{J}}_\ell^{(r)} = \mathbf{I}_{\prod_{n=1}^{r-1} M_n} \otimes \mathbf{J}_\ell^{(r)} \otimes \mathbf{I}_{\prod_{n=r+1}^R M_n}$, for $\ell = 1, 2$ and $r = 1, 2, \dots, R$, where $\mathbf{J}_\ell^{(r)} \in \mathbb{R}^{M_r^{(\text{sel})} \times M_r}$ are the selection matrices which select the $M_r^{(\text{sel})}$ elements belonging to the first and the second subarray in the r -th mode, respectively.

Since this expansion for R -D Standard ESPRIT is explicit in the perturbation of the subspace estimate and R -D Standard Tensor-ESPRIT only differs in the fact that it uses the enhanced HOSVD-based subspace estimate, we immediately conclude that a first order perturbation expansion for R -D Standard Tensor-ESPRIT is given by

$$\Delta\mu_k^{(r)} = \text{Im}\left\{ \mathbf{p}_k^T \cdot (\tilde{\mathbf{J}}_1^{(r)} \cdot \mathbf{U}_s)^+ \cdot [\tilde{\mathbf{J}}_2^{(r)} / \lambda_k^{(r)} - \tilde{\mathbf{J}}_1^{(r)}] \cdot [\Delta\hat{\mathbf{U}}^{[s]}]_{(R+1)}^T \cdot \mathbf{q}_k \right\} + \mathcal{O}\{\Delta^2\}. \quad (32)$$

An explicit expansion of $\Delta\mu_k^{(r)}$ in terms of the noise tensor \mathcal{N} is obtained by inserting the previous result (26).

C. Mean Square Errors

As it has been said above, the advantage of the first order perturbation expansion we have discussed so far is that it is explicit in the perturbation term \mathbf{N} and hence makes no assumptions about its distribution. However, it is often also desirable to know the mean square error if a specific distribution is assumed and the ensemble average over all possible noise realizations is computed.

In the sequel we show that the mean square error only depends on the second-order moments of the noise samples. Hence, we can derive the MSE as a function of the covariance matrix and the pseudo-covariance matrix only assuming the noise to be zero mean. We neither need to assume Gaussianity nor circular symmetry. For simplicity we consider the special case $R = 2$ for Standard Tensor-ESPRIT, however, a generalization to a larger number of dimensions is quite straightforward.

³If the shift invariance equations are solved completely independently, the correct pairing of the parameters across dimensions has to be found in a subsequent step. This is often avoided by computing the LS solutions for $\boldsymbol{\Psi}^{(r)}$ independently but then performing a joint eigendecomposition of all R dimensions to yield $\boldsymbol{\Phi}^{(r)}$. This step is not included in the performance analysis presented in this section, since no performance results on joint eigendecompositions are available and it appears to be a very difficult task. Moreover, this step has indeed no impact on the asymptotic estimation error of the spatial frequencies for high SNRs since the eigenvectors become asymptotically equal. As shown in [16], the impact of the perturbation of the eigenvectors is of second-order and can hence be ignored in a first-order perturbation analysis.

Theorem 3. Assume that the entries of the perturbation term \mathbf{N} or \mathcal{N} are zero mean random variables with finite second-order moments described by the covariance matrix $\mathbf{R}_{\text{nn}} = \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^H\}$ and the complementary covariance matrix $\mathbf{C}_{\text{nn}} = \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^T\}$ for $\mathbf{n} = \text{vec}\{\mathbf{N}\} = \text{vec}\{[\mathcal{N}]_{(3)}^T\}$. Then, the first-order approximation of the mean square estimation error for the k -th spatial frequency in the r -th mode is given by

$$\mathbb{E}\left\{\left(\Delta\mu_k^{(r)}\right)^2\right\} = \frac{1}{2}\left(\mathbf{r}_k^{(r)H} \cdot \mathbf{W}_{\text{mat}}^* \cdot \mathbf{R}_{\text{nn}}^T \cdot \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_k^{(r)} - \text{Re}\left\{\mathbf{r}_k^{(r)T} \cdot \mathbf{W}_{\text{mat}} \cdot \mathbf{C}_{\text{nn}} \cdot \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_k^{(r)}\right\}\right) + \mathcal{O}\left\{\text{Tr}(\mathbf{R}_{\text{nn}})^2\right\} \quad (33)$$

for R -D Standard ESPRIT and

$$\mathbb{E}\left\{\left(\Delta\mu_k^{(r)}\right)^2\right\} = \frac{1}{2}\left(\mathbf{r}_k^{(r)H} \cdot \mathbf{W}_{\text{ten}}^* \cdot \mathbf{R}_{\text{nn}}^T \cdot \mathbf{W}_{\text{ten}}^T \cdot \mathbf{r}_k^{(r)} - \text{Re}\left\{\mathbf{r}_k^{(r)T} \cdot \mathbf{W}_{\text{ten}} \cdot \mathbf{C}_{\text{nn}} \cdot \mathbf{W}_{\text{ten}}^T \cdot \mathbf{r}_k^{(r)}\right\}\right) + \mathcal{O}\left\{\text{Tr}(\mathbf{R}_{\text{nn}})^2\right\} \quad (34)$$

for 2-D Standard Tensor-ESPRIT⁴. The vector $\mathbf{r}_k^{(r)}$ and the matrices \mathbf{W}_{mat} and \mathbf{W}_{ten} are given by

$$\mathbf{r}_k^{(r)} = \mathbf{q}_k \otimes \left(\left[\left(\tilde{\mathbf{J}}_1^{(r)} \mathbf{U}_s \right)^+ \left(\tilde{\mathbf{J}}_2^{(r)} / e^{j\mu_k^{(r)}} - \tilde{\mathbf{J}}_1^{(r)} \right) \right]^T \cdot \mathbf{p}_k \right) \quad (35)$$

$$\begin{aligned} \mathbf{W}_{\text{mat}} &= \left(\Sigma_s^{-1} \cdot \mathbf{V}_s^T \right) \otimes \left(\mathbf{U}_n \cdot \mathbf{U}_n^H \right) \\ \mathbf{W}_{\text{ten}} &= \left(\Sigma_3^{[s]-1} \mathbf{U}_3^{[s]H} \right) \otimes \left([\mathbf{T}_1 \otimes \mathbf{T}_2] \mathbf{V}_3^{[n]*} \mathbf{V}_3^{[n]T} \right) \\ &+ \left(\mathbf{U}_s^T \otimes \mathbf{I}_M \right) \bar{\mathbf{T}}_2 \left(\mathbf{U}_1^{[s]*} \Sigma_1^{[s]-1} \mathbf{V}_1^{[s]T} \otimes \mathbf{U}_1^{[n]} \mathbf{U}_1^{[n]H} \right) \cdot \mathbf{K}_{M_2 \times (M_1 \cdot N)} \\ &+ \left(\mathbf{U}_s^T \otimes \mathbf{I}_M \right) \bar{\mathbf{T}}_1 \left(\mathbf{U}_2^{[s]*} \Sigma_2^{[s]-1} \mathbf{V}_2^{[s]T} \otimes \mathbf{U}_2^{[n]} \mathbf{U}_2^{[n]H} \right) \end{aligned} \quad (36)$$

$$\bar{\mathbf{T}}_1 = \begin{bmatrix} \mathbf{I}_{M_2} \otimes \mathbf{t}_{1,1} \\ \vdots \\ \mathbf{I}_{M_2} \otimes \mathbf{t}_{1,M_1} \end{bmatrix} \otimes \mathbf{I}_{M_2}, \bar{\mathbf{T}}_2 = \mathbf{I}_{M_1} \otimes \begin{bmatrix} \mathbf{I}_{M_1} \otimes \mathbf{t}_{2,1} \\ \vdots \\ \mathbf{I}_{M_1} \otimes \mathbf{t}_{2,M_2} \end{bmatrix},$$

and $\mathbf{t}_{r,m}$ is the m -th column of \mathbf{T}_r . Finally, $\mathbf{K}_{p,q}$ is the commutation matrix from (7).

Proof: cf. Appendix C.

Note that for 1-D Standard ESPRIT, this MSE expression agrees with the one shown in [18]. However, [18] does not directly generalize to the tensor case. This is the advantage of the MSE expressions (33) and (34) where we only need to replace \mathbf{W}_{mat} by \mathbf{W}_{ten} to account for the enhanced signal subspace estimate. Furthermore, note that the special case of circularly symmetric white noise corresponds to $\mathbf{R}_{\text{nn}} = \sigma_n^2 \cdot \mathbf{I}_{MN}$ and $\mathbf{C}_{\text{nn}} = \mathbf{0}_{MN \times MN}$.

D. Incorporation of Forward-Backward-Averaging

So far we have shown the explicit expansion and the MSE expressions for R -D Standard ESPRIT and R -D Standard Tensor-ESPRIT. In order to extend these results to Unitary-ESPRIT-type algorithms we need to incorporate the mandatory preprocessing for Unitary ESPRIT which is given by Forward-Backward-Averaging. The second step in Unitary ESPRIT is the transformation on the real-valued domain. However, it can

be shown that this step has no impact on the performance for high SNRs. Therefore, the asymptotic performance of Unitary-ESPRIT-type algorithms is found once Forward-Backward-Averaging is taken into account.

Forward-Backward-Averaging augments the N observations of the sampled R -D harmonics by N new “virtual” observations which are a conjugated and row- as well as column-flipped version of the original ones [9]. This can be expressed in matrix form as

$$\mathbf{X}^{(\text{fba})} = [\mathbf{X} \quad \mathbf{\Pi}_M \cdot \mathbf{X}^* \cdot \mathbf{\Pi}_N] \in \mathbb{C}^{M \times 2N} \quad (37)$$

Inserting $\mathbf{X} = \mathbf{X}_0 + \mathbf{N}$ we find

$$\begin{aligned} \mathbf{X}^{(\text{fba})} &= [\mathbf{X}_0, \quad \mathbf{\Pi}_M \cdot \mathbf{X}_0^* \cdot \mathbf{\Pi}_N] + [\mathbf{N}, \quad \mathbf{\Pi}_M \cdot \mathbf{N}^* \cdot \mathbf{\Pi}_N] \\ &= \mathbf{X}_0^{(\text{fba})} + \mathbf{N}^{(\text{fba})}. \end{aligned} \quad (38)$$

However, the latter relation shows that we are interested in the perturbation of the subspace of a matrix $\mathbf{X}_0^{(\text{fba})}$ superimposed by an additive perturbation $\mathbf{N}^{(\text{fba})}$, which is small. Since the explicit perturbation expansion we have used up to this point requires no additional assumptions, the surprisingly simple answer is that we do not need to change anything but we can apply the previous results directly. All we need to do is to replace all exact (noise-free) subspaces of \mathbf{X}_0 by the corresponding subspaces of $\mathbf{X}_0^{(\text{fba})}$. From (31), we immediately obtain the following explicit first-order expansion which is valid for R -D Unitary ESPRIT

$$\begin{aligned} \Delta\mu_k^{(r)} &= \text{Im}\left\{ \mathbf{p}_k^{(\text{fba})T} \cdot \left(\tilde{\mathbf{J}}_1^{(r)} \cdot \mathbf{U}_s^{(\text{fba})} \right)^+ \cdot \left[\tilde{\mathbf{J}}_2^{(r)} / \lambda_k^{(r)} - \tilde{\mathbf{J}}_1^{(r)} \right] \right. \\ &\quad \left. \cdot \Delta\mathbf{U}_s^{(\text{fba})} \cdot \mathbf{q}_k^{(\text{fba})} \right\} + \mathcal{O}\{\Delta^2\} \end{aligned} \quad (39)$$

where $\Delta\mathbf{U}_s^{(\text{fba})}$ is given by

$$\Delta\mathbf{U}_s^{(\text{fba})} = \mathbf{U}_n^{(\text{fba})} \cdot \mathbf{U}_n^{(\text{fba})H} \cdot \mathbf{N}^{(\text{fba})} \cdot \mathbf{V}_s^{(\text{fba})} \cdot \Sigma_s^{(\text{fba})^{-1}} \quad (40)$$

and $\mathbf{U}_s^{(\text{fba})}$, $\mathbf{U}_n^{(\text{fba})}$, $\mathbf{V}_s^{(\text{fba})}$, $\Sigma_s^{(\text{fba})}$ correspond to the signal subspace, the noise subspace, the row space, and the singular values of $\mathbf{X}_0^{(\text{fba})}$, respectively. Likewise, $\mathbf{q}_k^{(\text{fba})}$ and $\mathbf{p}_k^{(\text{fba})}$ represent the corresponding versions of \mathbf{q}_k and \mathbf{p}_k if \mathbf{U}_s is replaced by $\mathbf{U}_s^{(\text{fba})}$ in the shift invariance equations.

With the same reasoning, an explicit expansion for R -D Unitary Tensor-ESPRIT is obtained by consistently replacing \mathbf{X}_0 by $\mathbf{X}_0^{(\text{fba})}$ in (32), i.e.,

$$\begin{aligned} \Delta\mu_k^{(r)} &= \text{Im}\left\{ \mathbf{p}_k^{(\text{fba})T} \cdot \left(\tilde{\mathbf{J}}_1^{(r)} \cdot \mathbf{U}_s^{(\text{fba})} \right)^+ \cdot \left[\tilde{\mathbf{J}}_2^{(r)} / \lambda_k^{(r)} - \tilde{\mathbf{J}}_1^{(r)} \right] \right. \\ &\quad \left. \cdot \left[\Delta\hat{\mathcal{U}}^{[s](\text{fba})} \right]_{(R+1)}^T \cdot \mathbf{q}_k^{(\text{fba})} \right\} + \mathcal{O}\{\Delta^2\}. \end{aligned} \quad (41)$$

Similarly, Theorem 3 can be applied to compute the MSE since we only assumed the noise to be zero mean and possess finite second order moments, which is still true after forward-backward averaging. The following theorem summarizes the results for R -D Unitary ESPRIT and 2-D Unitary Tensor-ESPRIT:

Theorem 4. For the case where \mathbf{N} or \mathcal{N} contain zero mean random variables with finite second-order moments

⁴The reason that this result is specific for $R = 2$ is not (34) (which applies to arbitrary R) but (36) which we develop only for $R = 2$ here.

described by the covariance matrix $\mathbf{R}_{\text{nn}} = \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^H\}$ and the complementary covariance matrix $\mathbf{C}_{\text{nn}} = \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^T\}$ for $\mathbf{n} = \text{vec}\{\mathbf{N}\} = \text{vec}\left\{[\mathcal{N}]_{(3)}^T\right\}$, the MSE for R -D Unitary ESPRIT and 2-D Unitary Tensor-ESPRIT are given by (33) and (34) if we replace $\mathbf{r}_k^{(r)}$ by $\mathbf{r}_k^{(r)(\text{fba})}$, \mathbf{W}_{mat} by $\mathbf{W}_{\text{mat}}^{(\text{fba})}$, \mathbf{W}_{ten} by $\mathbf{W}_{\text{ten}}^{(\text{fba})}$, and \mathbf{R}_{nn} as well as \mathbf{C}_{nn} by $\mathbf{R}_{\text{nn}}^{(\text{fba})}$ and $\mathbf{C}_{\text{nn}}^{(\text{fba})}$. Here, $\mathbf{r}_k^{(r)(\text{fba})}$, $\mathbf{W}_{\text{mat}}^{(\text{fba})}$, and $\mathbf{W}_{\text{ten}}^{(\text{fba})}$ are computed as in (33) and (34) by consistently replacing all quantities by their forward-backward-averaged equivalents. Moreover, $\mathbf{R}_{\text{nn}}^{(\text{fba})}$ and $\mathbf{C}_{\text{nn}}^{(\text{fba})}$ represent the covariance and the pseudo-covariance matrix of the forward-backward averaged noise, which are given by

$$\begin{aligned} \mathbf{R}_{\text{nn}}^{(\text{fba})} &= \begin{bmatrix} \mathbf{R}_{\text{nn}} & \mathbf{C}_{\text{nn}} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbf{C}_{\text{nn}}^* & \mathbf{\Pi}_{MN} \cdot \mathbf{R}_{\text{nn}}^* \cdot \mathbf{\Pi}_{MN} \end{bmatrix} \\ \mathbf{C}_{\text{nn}}^{(\text{fba})} &= \begin{bmatrix} \mathbf{C}_{\text{nn}} & \mathbf{R}_{\text{nn}} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbf{R}_{\text{nn}}^* & \mathbf{\Pi}_{MN} \cdot \mathbf{C}_{\text{nn}}^* \cdot \mathbf{\Pi}_{MN} \end{bmatrix}. \end{aligned}$$

Proof: cf. Appendix D.

Note that in the special case where the noise is circularly symmetric and white we have $\mathbf{R}_{\text{nn}}^{(\text{fba})} = \sigma_n^2 \cdot \mathbf{I}_{2MN}$ and $\mathbf{C}_{\text{nn}}^{(\text{fba})} = \sigma_n^2 \cdot \mathbf{\Pi}_{2MN}$.

It is important to note that results on Unitary ESPRIT in this section relate to the LS solution only. If TLS is used instead, the equivalence of Standard ESPRIT with Forward-Backward-Averaging and Unitary ESPRIT is shown in [9].

E. Extension to other ESPRIT-type algorithms

In a similar manner as in the previous section, other ESPRIT-type algorithms can be analyzed. For instance, the NC Standard ESPRIT and NC Unitary ESPRIT algorithm for strict-sense non-circular sources are based on a different kind of preprocessing where instead of augmenting the columns we augment the rows of the measurement matrix. Yet, the explicit first order perturbation expansion still applies since the result can be written as a noise-free (augmented) measurement matrix superimposed by a small (augmented) perturbation matrix. Consequently, for the explicit expansion we only need to consistently replace the quantities originating from the SVD of \mathbf{X}_0 by the corresponding quantities from the appropriately preprocessed measurement matrix $\mathbf{X}_0^{(\text{nc})}$. Likewise, the MSE expressions are directly applicable since we only require the noise to be zero mean and possess finite second-order moments.

Another possible extension is to incorporate spatial smoothing. If sources are mutually coherent, preprocessing must be applied to the data to decorrelate the sources prior to any subspace-based parameter estimation scheme. Via Forward-Backward-Averaging, two sources can be decorrelated. However, if more than two sources are coherent (or if FBA cannot be applied), additional preprocessing is needed. For spatial smoothing we divide the array into a number of identical displaced subarrays and average the spatial covariance matrix over these subarrays. Since the number of subarrays we choose is a design parameter influencing the performance, investigating its effect by virtue of an analytical performance assessment would be desirable. Note that the spatial averaging

introduces a correlation into the noise. Therefore, the presented framework is particularly attractive since for the explicit expansion, no assumptions about the noise statistics are needed. A further extension is the performance assessment of tensor-based schemes for spatial smoothing. We have introduced a tensor-based formulation of spatial smoothing for R -D signals in [11]. Moreover, a tensor-based spatial smoothing technique for 1-D damped and undamped harmonic retrieval with a single snapshot is shown in [38]. The extension to multiple snapshots is introduced in [39] and an R -D extension is shown in [37]. A major advantage of [39], [37] is that the performance of the ESPRIT-type parameter estimates is almost independent of the choice of the subarray size. This could be verified by analytical results if the performance analysis is extended accordingly.

F. Incorporation of Structured Least Squares (SLS)

So far, all performance results are based on ESPRIT using LS, i.e., the overdetermined shift invariance equations are solved using LS only. However, the LS solution to the shift invariance equation is in general suboptimal as errors on both sides of the equations need to be taken into account. Even more so, since for overlapping subarrays, the shift invariance equation has a specific structure resulting in common error terms on both sides of the equations, this structure should be taken into account when solving them. This has led to the development of the SLS algorithm [8]. Since it has been shown that the resulting ESPRIT algorithm using SLS outperforms ESPRIT using LS and TLS for overlapping subarrays [8], it is desirable to extend our performance analysis results to SLS-based ESPRIT as well.

Due to the fact that our analysis is asymptotic in the SNR we can make the following simplifying assumptions for SLS. Firstly, we consider only a single iteration, as proposed in [8]. This is optimal for high SNRs, since the underlying cost function is quadratic but actually asymptotically linear (the quadratic term vanishes against the linear terms for high SNRs). Secondly, we do not consider the optional regularization term in SLS (i.e., we set the corresponding regularization parameter α to infinity) as regularization is typically not needed for high SNRs.

Under these conditions we can show the following theorem:

Theorem 5. A first order expansion of the estimation error of 1-D Standard ESPRIT using SLS is given by

$$\Delta\mu_{k,\text{SLS}} = \text{Im}\{\mathbf{r}_{k,\text{SLS}}^T \cdot \text{vec}\{\Delta\mathbf{U}_s\}\} + \mathcal{O}\{\Delta^2\} \quad (42)$$

$$= \text{Im}\{\mathbf{r}_{k,\text{SLS}}^T \cdot \mathbf{W}_{\text{mat}} \cdot \text{vec}\{\mathbf{N}\}\} + \mathcal{O}\{\Delta^2\} \quad (43)$$

where \mathbf{W}_{mat} is defined in (35) and $\mathbf{r}_{k,\text{SLS}}^T$ is given by

$$\begin{aligned} \mathbf{r}_{k,\text{SLS}}^T &= \mathbf{q}_k^T \otimes \left[\mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \left(\frac{\mathbf{J}_2}{e^{j\mu_k}} - \mathbf{J}_1 \right) \right] \\ &\quad - \left(\mathbf{q}_k^T \otimes \left[\mathbf{p}_k^T \cdot \frac{(\mathbf{J}_1 \cdot \mathbf{U}_s)^H}{e^{j\mu_k}} \right] \right) \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1} \cdot \mathbf{W}_{\text{R,U}} \\ \mathbf{W}_{\text{R,U}} &= (\mathbf{\Psi}^T \otimes \mathbf{J}_1) + \mathbf{I}_d \otimes \left(\mathbf{J}_1 \cdot \mathbf{U}_s (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_2 \right) \\ &\quad - \mathbf{\Psi}^T \otimes \left(\mathbf{J}_1 \cdot \mathbf{U}_s (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_1 \right) - (\mathbf{I}_d \otimes \mathbf{J}_2) \end{aligned}$$

$$\mathbf{F}_{\text{SLS}} = \left[\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \mathbf{U}_s), \left(\Psi^T \otimes \mathbf{J}_1 \right) - (\mathbf{I}_d \otimes \mathbf{J}_2) \right]$$

for $k = 1, 2, \dots, d$. The MSE for zero mean noise samples can then be computed via

$$\begin{aligned} \mathbb{E} \left\{ (\Delta \mu_{k,\text{SLS}})^2 \right\} &= \frac{1}{2} \left(\mathbf{r}_{k,\text{SLS}}^H \cdot \mathbf{W}_{\text{mat}}^* \cdot \mathbf{R}_{\text{nn}}^T \cdot \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_{k,\text{SLS}} - \right. \\ &\left. \text{Re} \left\{ \mathbf{r}_{k,\text{SLS}}^T \cdot \mathbf{W}_{\text{mat}} \cdot \mathbf{C}_{\text{nn}} \cdot \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_{k,\text{SLS}} \right\} \right) + \mathcal{O} \left\{ \text{Tr}(\mathbf{R}_{\text{nn}})^2 \right\} \end{aligned} \quad (44)$$

Proof: Equation (43) is shown in Appendix E. Since the explicit expansion of (43) has the same form as the explicit expansion in (30), the MSE expression (44) is shown analogously to Theorem 3 as presented in Appendix C.

G. Special case: Single source

So far we have found closed-form expressions for the first-order approximate MSE for different kinds of ESPRIT-type algorithms. As they are deterministic, they can be plotted for varying system parameters without performing Monte-Carlo simulations and one can learn from these plots under which conditions the performance changes how much.

However, it would be desirable to find expressions that are even more insightful. The biggest disadvantage of the MSE expressions in their current form is that they are formulated in terms of the subspaces of the noise-free observation matrix and not in terms of the actual parameters with a physical significance, such as, the number of sensors or the positions of the sources.

Finding such a formulation in the general case seems to be impossible given the complicated algebraic nature in which the MSE expressions depend on the physical parameters. However, it becomes much easier if some special cases are considered. Therefore we present one example of such a special case in this section, namely, the case of a single source captured by a uniform linear array (ULA) and a uniform rectangular array (URA) and circularly symmetric white noise. Although this is a very trivial case, it serves as an example which types of insights such an analytical performance assessment can provide. For the 1-D case we have the following theorem:

Theorem 6. *For the case of an M -element ULA (1-D) and a single source ($d = 1$) we can show that the mean square estimation error of the spatial frequency for Standard ESPRIT and for Unitary ESPRIT is given by*

$$\mathbb{E} \left\{ (\Delta \mu)^2 \right\} = \frac{1}{\hat{\rho}} \cdot \frac{1}{(M-1)^2} + \mathcal{O} \left\{ \frac{1}{\hat{\rho}^2} \right\} \quad (45)$$

Moreover, the deterministic Cramér-Rao Bound can be simplified into

$$\text{CRB} = \frac{1}{\hat{\rho}} \cdot \frac{6}{M \cdot (M^2 - 1)} \quad (46)$$

Consequently, the asymptotic efficiency is given by

$$\eta = \lim_{\hat{\rho} \rightarrow \infty} \frac{\text{CRB}}{\mathbb{E} \left\{ (\Delta \mu)^2 \right\}} = \frac{6(M-1)}{M(M+1)}. \quad (47)$$

Here, $\hat{\rho}$ represents the effective SNR given by $\hat{\rho} = \frac{\hat{P}_T \cdot N}{\sigma_n^2}$, where \hat{P}_T is the empirical transmit power given by $\hat{P}_T = \|\mathbf{S}\|_F^2 / N$ if \mathbf{S} is the matrix of source symbols.

Proof: cf. Appendix F.

Note that [28] provide an MSE expression for ESPRIT for the case of a single source which scales with $1/M^2$ and is derived under the assumption of high “array SNR” $P \cdot M / \sigma_n^2$, i.e., it is asymptotic also in M . The result presented here is accurate for small values of M as well and only asymptotic in the effective SNR $N \cdot P_T / \sigma_n^2$. Also note that analytical expression for the stochastic Cramér-Rao Bound for one and two sources are available in [33].

We can simplify the MSE expression for ESPRIT using SLS shown in Section V-F in a similar manner, as shown in the following theorem:

Theorem 7. *The mean square estimation error of the spatial frequency for Standard ESPRIT using SLS on an M -element ULA is given by*

$$\mathbb{E} \left\{ (\Delta \mu)^2 \right\} = \frac{6}{\hat{\rho}} \cdot \frac{M^4 - 2M^3 + 24M^2 - 22M + 23}{M(M^2 + 11)^2(M-1)^2} + \mathcal{O} \left\{ \frac{1}{\hat{\rho}^2} \right\} \quad (48)$$

Consequently, the asymptotic efficiency is given by

$$\begin{aligned} \eta &= \frac{(M^2 + 11)^2(M-1)}{(M+1)(M^4 - 2M^3 + 24M^2 - 22M + 23)} \\ &= \frac{M^5 - M^4 + 22M^3 - 22M^2 + 121M - 121}{M^5 - M^4 + 22M^3 + 2M^2 + M + 23} \end{aligned} \quad (49)$$

Proof: cf. Appendix G.

Finally, for the 2-D case, we have the following result:

Theorem 8. *For a uniform rectangular array of $M_1 \times M_2$ sensors (2-D) and a single source, the mean square estimation error of the spatial frequency for 2-D Standard ESPRIT, 2-D Standard Tensor-ESPRIT, 2-D Unitary ESPRIT, and 2-D Unitary Tensor-ESPRIT can be simplified into*

$$\begin{aligned} \mathbb{E} \left\{ (\Delta \mu^{(1)})^2 + (\Delta \mu^{(2)})^2 \right\} \\ = \frac{1}{\hat{\rho}} \cdot \left(\frac{1}{(M_1 - 1)^2 \cdot M_2} + \frac{1}{M_1 \cdot (M_2 - 1)^2} \right) + \mathcal{O} \left\{ \frac{1}{\hat{\rho}^2} \right\} \end{aligned} \quad (50)$$

Finally, the deterministic Cramér-Rao Bound for a URA can be written as

$$\text{CRB} = \text{Tr}(\mathbf{C}) = \frac{1}{\hat{\rho}} \cdot \left(\frac{6}{M \cdot (M_1^2 - 1)} + \frac{6}{M \cdot (M_2^2 - 1)} \right). \quad (51)$$

Proof: in Appendix H we derive MSE expressions for R -D Standard ESPRIT, R -D Unitary ESPRIT, and the Cramér-Rao Bound for the more general R -D case. From these, this theorem follows by setting $R = 2$. Moreover, we show the identity of R -D Standard Tensor-ESPRIT and R -D Unitary Tensor-ESPRIT with R -D Standard ESPRIT for $R = 2$.

These MSE expressions provide some interesting insights. Firstly, they show that for a single source there is neither an improvement in terms of the estimation accuracy from applying Forward-Backward-Averaging nor from the HOSVD-based subspace estimate. This is surprising at first sight since

the HOSVD-based subspace estimate itself is more accurate also for a single source.

Moreover, they show that the asymptotic efficiency can be explicitly computed and that it is only a function of the array geometry, i.e., the number of sensors in the array. Unfortunately, the outcome of this analysis is that ESPRIT-type algorithms using LS are asymptotically efficient for $M = 2, 3$ in the 1-D case and $M_1 \in [2, 3]$, $M_2 \in [2, 3]$ in the 2-D case. However, they become less and less efficient when the number of sensors grows, in fact, for $M \rightarrow \infty$ we even have $\eta \rightarrow 0$. A possible explanation for this phenomenon could be that an M -sensor ULA offers not only the one shift invariance used in LS (the first and last $M - 1$ sensors) but multiple invariances [36], which are not fully exploited by LS. However, for ESPRIT based on SLS, the asymptotic efficiency is in general higher, in fact, we have $\eta \rightarrow 1$ for $M = 2$, $M = 3$ and $M \rightarrow \infty$ for a single source. Moreover, even for limited M , η is never far away from 1. As we show in the simulations below, we have $\eta = 1$ for $M = 2, 3$ and the smallest value of η is obtained for $M = 5$ where $\eta = 36/37 \approx 0.973$ for $d = 1$.

VI. SIMULATION RESULTS

In this section we show numerical results to demonstrate the asymptotic behavior of the analytical performance assessment presented in this chapter. We first investigate the subspace estimation accuracy in order to verify (26). Note that the analytical results for the subspace estimates are explicit expansions in terms of the perturbation (i.e., the additive noise). Therefore, we repeat the experiment with a number of randomly generated realizations of the noise and perform Monte-Carlo averaging over the analytical expansions. These “**semi-analytical**” results are then compared with purely **empirical** results where we estimate the subspace via an SVD or a HOSVD and compute the estimation error compared to the true signal subspace.

The subsequent numerical results demonstrate the performance of ESPRIT-type parameter estimation schemes. Here, we compute the mean square estimation error in three different ways. Firstly, **analytically**, via the MSE expressions provided in Theorem 3, Theorem 4, and Theorem 5 (eqn. (44)), respectively. Secondly, **semi-analytically**, by performing Monte-Carlo averaging over the explicit first-order expansions provided in equation (32), (41), and (43), respectively. Thirdly, **empirically**, by estimating the spatial frequencies via the corresponding ESPRIT-type algorithms and comparing the estimates to the true spatial frequencies.

For all the simulations we assume a known number of planar wavefronts impinging on an antenna array of M isotropic sensor elements. We assume uniform $\lambda/2$ spacing in all dimensions, i.e., an M -element uniform linear array (ULA) in the 1-D case and an $M_1 \times M_2$ uniform rectangular array (URA) in the 2-D case. The sources emit narrow-band waveforms $s_i(t)$ modeled as complex Gaussian distributed symbols $s_i(t)$ and we observe N subsequent snapshots $t = 1, 2, \dots, N$. All sources are assumed to have unit power, i.e., $\mathbb{E}\{|s_i(t)|^2\} = 1$. In the case where source correlation is investigated we generate the symbols $s_i(t)$ such that $\mathbb{E}\{s_i(t) \cdot s_j(t)^*\} = \rho \cdot e^{j\varphi_{i,j}}$ for

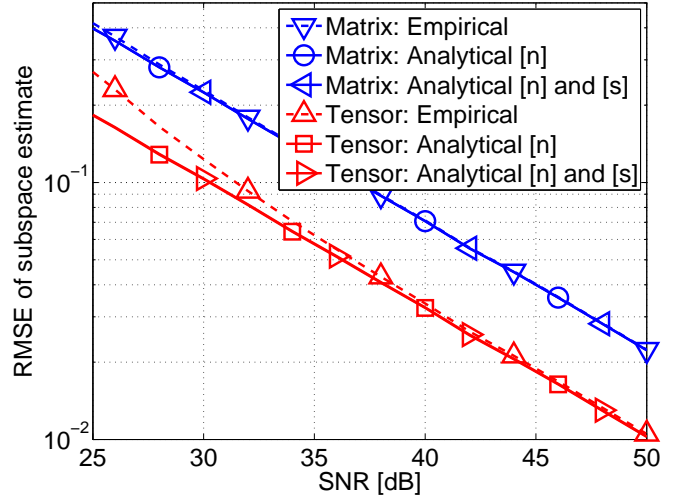


Figure 1. Subspace estimation accuracy using $\Gamma^{[n]}$ only vs. using $\Gamma^{[n]}$ and $\Gamma^{[s]}$. Scenario: $d = 3$ correlated sources ($\rho = 0.97$) at $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$, an 8×8 URA, and $N = 20$ snapshots.

$i \neq j = 1, 2, \dots, d$, where ρ is the correlation coefficient between each pair of sources and $\varphi_{i,j}$ is a uniformly distributed correlation phase. The additive noise is generated according to a circularly symmetric complex Gaussian distribution with zero mean and variance σ_n^2 and noise samples are assumed to be mutually independent. Therefore, the Signal to Noise Ratio (SNR) is defined as $1/\sigma_n^2$.

A. Subspace estimation accuracy

We evaluate the subspace estimation accuracy by computing the Frobenius norm of the subspace estimation error, i.e., $\|\Delta \mathbf{U}_s\|_F^2$ in the matrix case and $\left\| \left[\Delta \mathcal{U}^{[s]} \right]_{(R+1)}^T \right\|_H^2$ in the tensor case.

In order to find the estimation error empirically, we obtain a subspace estimate $\hat{\mathbf{U}}_s$ via an SVD of the noisy observation and then compare it to the true subspace \mathbf{U}_s column by column. The estimation error of the n -th column is computed via

$$\Delta \mathbf{u}_n = \hat{\mathbf{u}}_n \cdot \frac{\hat{\mathbf{u}}_n^H \mathbf{u}_n}{|\hat{\mathbf{u}}_n^H \mathbf{u}_n|} - \mathbf{u}_n, \quad n = 1, 2, \dots, d \quad (52)$$

to account for the inherent phase ambiguity in each column of the SVD, cf. [19].

For the analytical estimation error we calculate $\Delta \mathbf{U}_s$ via the first-order expansion $\Delta \mathbf{U}_s \approx \mathbf{U}_n \cdot \Gamma^{[n]}$ provided in (19) and the expansion $\Delta \mathbf{U}_s \approx \mathbf{U}_n \cdot \Gamma^{[n]} + \mathbf{U}_s \cdot \Gamma^{[s]}$ provided in (20), respectively. Note that the latter is more accurate since it additionally considers the perturbation of the individual singular vectors, i.e., the particular choice of the basis for the signal subspace. However, this contribution is irrelevant for the performance of ESPRIT-type algorithms.

In Figure 1 we have $d = 3$ sources positioned at $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$ and mutually correlated with a correlation coefficient of

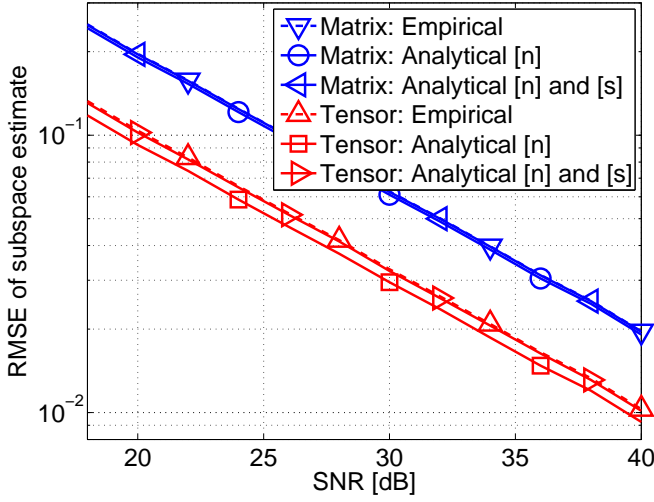


Figure 2. Subspace estimation accuracy using $\Gamma^{[n]}$ only vs. using $\Gamma^{[n]}$ and $\Gamma^{[s]}$. Scenario: $d = 4$ uncorrelated sources at $\mu_1^{(1)} = -1.5, \mu_2^{(1)} = 0.5, \mu_3^{(1)} = 1.0, \mu_4^{(1)} = -0.3, \mu_1^{(2)} = 1.3, \mu_2^{(2)} = -0.2, \mu_3^{(2)} = 0.7, \mu_4^{(2)} = -1.5$, an 8×8 URA, and $N = 5$ snapshots.

$\rho = 0.97$. Moreover, the array size is increased to an 8×8 URA. For the simulation result shown in Figure 2 we consider $d = 4$ uncorrelated sources located at $\mu_1^{(1)} = -1.5, \mu_2^{(1)} = 0.5, \mu_3^{(1)} = 1.0, \mu_4^{(1)} = -0.3, \mu_1^{(2)} = 1.3, \mu_2^{(2)} = -0.2, \mu_3^{(2)} = 0.7, \mu_4^{(2)} = -1.5$ and $N = 5$ snapshots.

Both simulations show that the empirical estimation errors agree with the analytical results as the SNR tends to infinity. Therefore, the improvement obtained by the HOSVD-based subspace estimate can reliably be predicted via the analytical expressions. In general, it is particularly pronounced for correlated sources and for a small number of snapshots. Moreover, while for three correlated sources, the impact of the additional term $U_s \cdot \Gamma^{[s]}$ is negligibly small, it is clearly visible for the four uncorrelated sources shown in Figure 2.

B. R-D Tensor-ESPRIT

The following set of simulation results demonstrates the performance of R -D matrix-based and tensor-based ESPRIT. As explained in the beginning of this section, for the analytical results we use (33) and (34) for R -D Standard ESPRIT and R -D Standard Tensor-ESPRIT, respectively, and their extensions for Unitary ESPRIT and R -D Unitary Tensor-ESPRIT as discussed in Theorem 4. Likewise, the semi-analytical results are obtained by Monte-Carlo averaging of the explicit first-order expansion provided in (32) and (41), respectively.

For Figure 3 we employ a 5×6 URA and collect $N = 20$ snapshots from two sources located at $\mu_1^{(1)} = 1, \mu_2^{(1)} = -0.5, \mu_1^{(2)} = -0.5, \mu_2^{(2)} = 1$. The sources are highly correlated with a correlation of $\rho = 0.9999$. On the other hand, for Figure 4 we increase the number of sources to $d = 3$ and the correlation coefficient to $\rho = 0.97$. Moreover, the spatial frequencies of the sources are given by $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$ and we use an 8×8 URA.

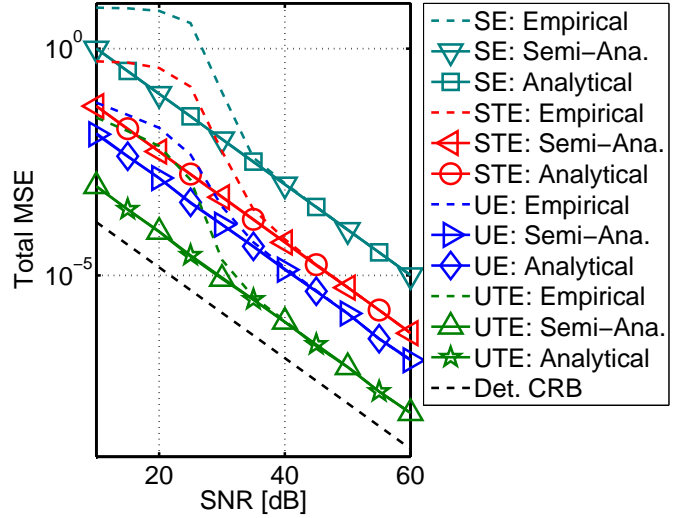


Figure 3. Performance of 2-D SE, STE, UE, UTE for $d = 2$ highly correlated sources ($\rho = 0.9999$) located at $\mu_1^{(1)} = 1, \mu_2^{(1)} = -0.5, \mu_1^{(2)} = -0.5, \mu_2^{(2)} = 1$, a 5×6 URA, and $N = 20$ snapshots.

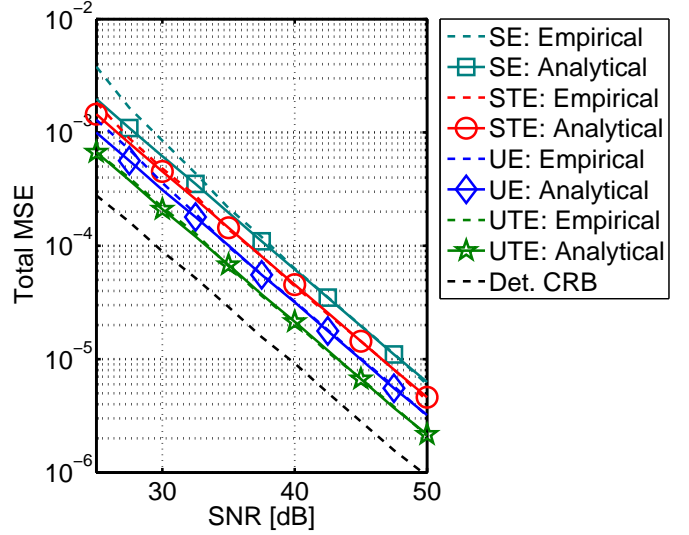


Figure 4. Performance of 2-D SE, STE, UE, UTE for $d = 3$ correlated sources ($\rho = 0.97$) positioned at $\mu_1^{(1)} = 0.7, \mu_2^{(1)} = 0.9, \mu_3^{(1)} = 1.1, \mu_1^{(2)} = -0.1, \mu_2^{(2)} = -0.3, \mu_3^{(2)} = -0.5$, an 8×8 URA, and $N = 20$ snapshots.

To enhance the legibility, we show the semi-analytical estimation errors only in Figure 3 since they always agree with the analytical results, as expected. Moreover, the empirical estimation errors agree with the analytical ones for high SNRs. This is also expected as the performance analysis framework presented here is asymptotically accurate for high effective SNRs. We conclude that the improvement in terms of estimation accuracy for Tensor-ESPRIT-type parameter estimation schemes can reliably be predicted via the analytical expressions we have derived.

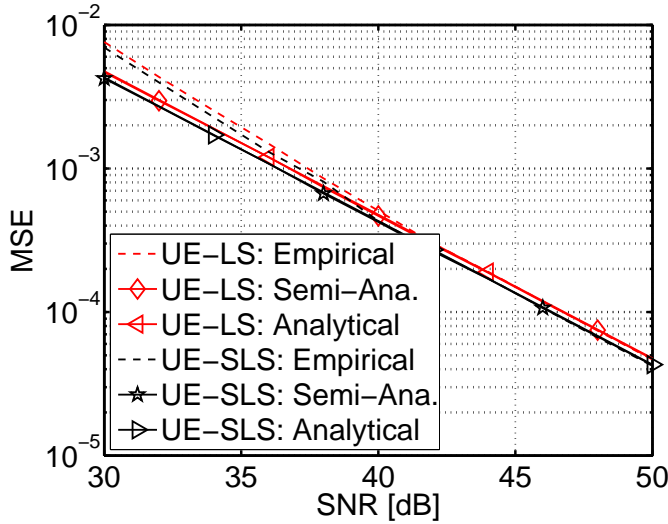


Figure 5. Performance of LS-ESPRIT vs. SLS-ESPRIT for 4 sources at $\mu_1 = 1.0, \mu_2 = 0.7, \mu_3 = -0.6, \mu_4 = -0.3$, an $M = 8$ ULA, $N = 3$ snapshots.

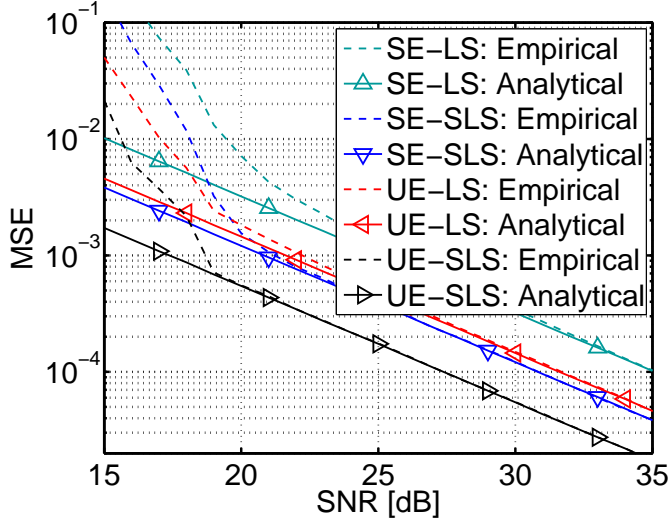


Figure 6. Performance of LS-ESPRIT vs. SLS-ESPRIT for $d = 3$ correlated sources ($\rho = 0.99$) at $\mu_1 = 1, \mu_2 = 0, \mu_3 = -1$, a $M = 12$ ULA and $N = 10$ snapshots.

C. Structured Least Squares

The next set of simulation results illustrates the analytical expressions for ESPRIT using SLS. The semi-analytical MSE is obtained by Monte-Carlo averaging over the explicit expansion provided in (43) and the analytical MSE is computed via (44). For the empirical estimation errors we perform a single iteration of the Structured Least Squares algorithm and do not use regularization (i.e., the regularization parameter α is set to ∞).

The first simulation result is shown in Figure 5. Here we consider $N = 3$ snapshots from $d = 4$ uncorrelated sources captured by an $M = 8$ element uniform linear array. The sources' spatial frequencies are given by $\mu_1 = 1.0, \mu_2 = 0.7, \mu_3 = -0.6, \mu_4 = -0.3$. Note that since $N < d$, we cannot

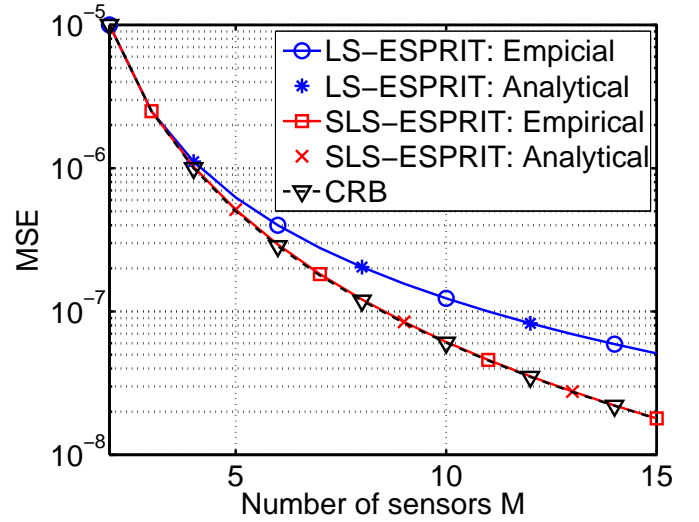


Figure 7. Performance of LS-ESPRIT and SLS-ESPRIT for a single source vs. the number of sensors M (M -ULA) at an effective SNR of 25 dB ($P_T = 1, \sigma_n^2 = 0.032, N = 10$).

apply Standard ESPRIT, therefore, only Unitary ESPRIT is used. On the other hand, in the second scenario we consider $N = 10$ snapshots from $d = 3$ sources that are mutually correlated with a correlation coefficient of $\rho = 0.99$. The sources are located at $\mu_1 = 1, \mu_2 = 0, \mu_3 = -1$ and a $M = 12$ element ULA is used. The corresponding estimation errors are shown in Figure 6.

As before, the empirical results agree with the analytical results for high SNRs. Moreover, the improvement in MSE obtained via SLS is particularly pronounced for the correlated sources. However, even the very slight improvement which is present for four uncorrelated sources can reliably be predicted via the analytical MSE expressions we have derived.

D. Asymptotic efficiency for a single source

The final set of simulation results demonstrates the special case of a single source, in which case the MSE expressions can be simplified to very compact closed-form expressions which only depend on the physical parameters, i.e., the array size and the SNR.

Figures 7 and 8 show the MSE vs. the number of sensors M for a (1-D) Uniform Linear Array and vs. M_1 for a (2-D) $M_1 \times M_1$ Uniform Rectangular Array, respectively. For both scenarios, the spatial frequencies of the single source were drawn randomly (note that they have no impact on the MSE). The effective SNR was set to 25 dB for Figure 7 ($P_T = 1, \sigma_n^2 = 0.032, N = 10$) and to 46 dB for Figure 8 ($P_T = 1, \sigma_n^2 = 10^{-4}, N = 4$), respectively.

For both plots we observe that LS-ESPRIT is asymptotically efficient for $M = 2$ and $M = 3$ (which, in the 2-D case means, a 3×3 URA) and then becomes increasingly inefficient as the array size grows. Moreover, for the 1-D case we see that SLS-based ESPRIT is in fact very close to the Cramér-Rao Bound, which may, at first sight, lead to believe that the asymptotic efficiency is in fact 1 for all M . However, as we have shown

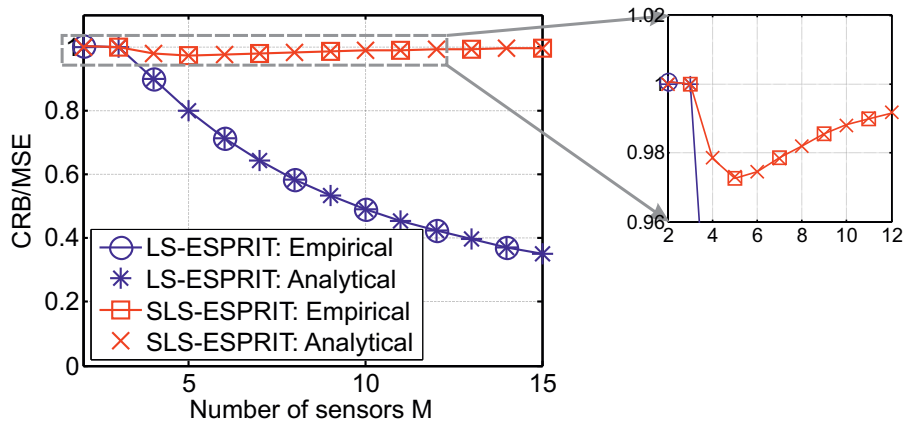


Figure 9. Asymptotic efficiency of LS-ESPRIT vs. SLS-ESPRIT vs. M . Same scenario as in Figure 7. The left-hand side shows a zoom.

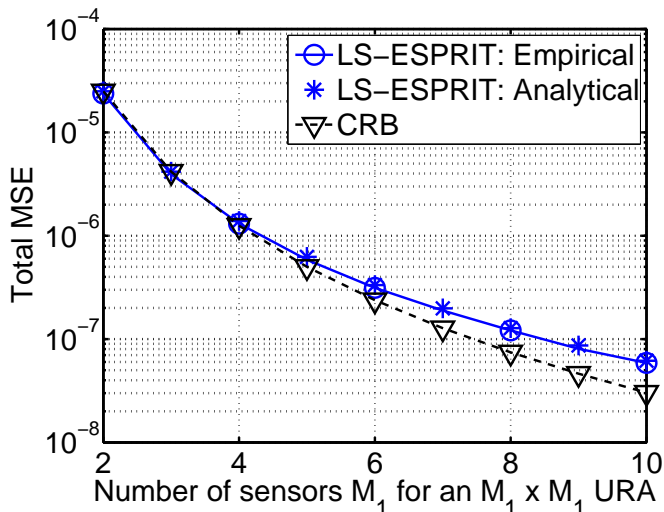


Figure 8. Performance of LS-ESPRIT for a single source vs. M_1 using an $M_1 \times M_1$ URA at an effective SNR of 46 dB ($P_T = 1$, $\sigma_n^2 = 10^{-4}$, $N = 4$).

it is in fact slightly lower than one. Therefore, we provide two additional figures where we depict the “asymptotic efficiency”, i.e., we divide the CRB by the corresponding value of the MSE. The resulting efficiency plot is shown in Figure 9. This plot shows more clearly that LS-ESPRIT becomes increasingly inefficient for $M > 3$, whereas SLS-ESPRIT approaches $\eta = 1$ for large M . The worst efficiency is found for $M = 5$ where we have $\eta = 36/37 \approx 0.973$.

VII. CONCLUSIONS

In this paper we have discussed a framework for analytical performance assessment of subspace-based parameter estimation schemes. It is based on earlier results on an explicit first-order expansion of the SVD and its application to 1-D versions of subspace-based parameter estimation schemes, e.g., ESPRIT. We have extended this framework in a number of ways. Firstly, we have derived an explicit first-order expansion of the HOSVD-based subspace estimate which is the basis for Tensor-ESPRIT-type algorithms. Secondly, we have shown that the first-order expansion for 1-D Standard ESPRIT can

be extended to other ESPRIT-type algorithms, e.g., R -D Standard ESPRIT, R -D Unitary ESPRIT, R -D Standard Tensor-ESPRIT, or R -D Unitary Tensor-ESPRIT. Thirdly, we have derived a corresponding first-order expansion for Structured Least Squared (SLS)-based ESPRIT-type algorithms.

All these expansions have in common that they are explicit, i.e., no assumption about the statistics of either desired signal or additive perturbation need to be made. We only require the perturbation to be small compared to the desired signal. We also do not need the number of snapshots to be large, i.e., they even apply to the single snapshot case ($N = 1$). Our fourth contribution is to show that the mean square error can readily be computed in closed-form and that it depends only on the second-order moments of the noise. Consequently, for the MSE expressions we only need the noise to be zero mean and its second order moments to be finite. Neither Gaussianity nor circular symmetry is required. This is a particularly attractive feature of our approach with respect to different types of preprocessing which alters the noise statistics, e.g., spatial smoothing (which yields spatially correlated noise) or forward-backward averaging (which annihilates the circular symmetry of the noise). Since we do not require spatial whiteness or circular symmetry, our MSE expressions are directly applicable. The resulting MSE expressions are asymptotic in the effective SNR, i.e., they become accurate as either the noise variance goes to zero or the number of observations goes to infinity.

As a final contribution we have investigated the special case of a single source, circularly symmetric white noise, and uniform linear (1-D) or uniform rectangular (2-D) arrays. In this case we have been able to show analytically, that R -D Standard ESPRIT, R -D Unitary ESPRIT, and (for $R = 2$) R -D Standard Tensor-ESPRIT as well as R -D Unitary Tensor-ESPRIT yield the same MSE, which only depends on the effective SNR and the number of antenna elements. We have also shown that 1-D Standard ESPRIT using SLS has a lower MSE which is also expressed explicitly as a function of the effective SNR and the number of antenna elements.

APPENDIX A
PROOF OF THEOREM 1

As shown in (16), the estimated signal subspace tensor can be computed via

$$\hat{\mathcal{U}}^{[s]} = \hat{\mathcal{S}}^{[s]} \times_1 \hat{\mathcal{U}}_1^{[s]} \dots \times_R \hat{\mathcal{U}}_R^{[s]} \times_{R+1} \hat{\Sigma}_s^{-1}. \quad (53)$$

Here, $\hat{\mathcal{S}}^{[s]}$ represents the truncated version core tensor $\hat{\mathcal{S}}$ from the HOSVD of \mathcal{X} . In order to eliminate $\hat{\mathcal{S}}^{[s]}$ we require the following Lemma:

Lemma 1. *The truncated core tensor $\hat{\mathcal{S}}^{[s]}$ can be computed from \mathcal{X} directly via*

$$\hat{\mathcal{S}} = \mathcal{X} \times_1 \hat{\mathcal{U}}_1^{[s]H} \dots \times_{R+1} \hat{\mathcal{U}}_{R+1}^{[s]H}. \quad (54)$$

Proof: To show (54) we insert the HOSVD of \mathcal{X} given by $\mathcal{X} = \hat{\mathcal{S}} \times_1 \hat{\mathcal{U}}_1 \dots \times_{R+1} \hat{\mathcal{U}}_{R+1}$. Using (4), we obtain

$$\hat{\mathcal{S}} = \mathcal{S} \times_1 \left(\hat{\mathcal{U}}_1^{[s]H} \cdot \hat{\mathcal{U}}_1 \right) \dots \times_{R+1} \left(\hat{\mathcal{U}}_{R+1}^{[s]H} \cdot \hat{\mathcal{U}}_{R+1} \right). \quad (55)$$

However, since the matrices of r -mode singular vectors $\hat{\mathcal{U}}_r$ are unitary, they satisfy $\hat{\mathcal{U}}_r^{[s]H} \cdot \hat{\mathcal{U}}_r = [\mathbf{I}_{p_r}, \mathbf{0}_{p_r \times (M_r - p_r)}]$. Therefore, $\hat{\mathcal{S}}^{[s]}$ computed via (54) contains the first p_r elements of $\hat{\mathcal{S}}$ in the r -th mode, which shows that it is indeed the truncated core tensor. ■

Next, we use Lemma 1 to eliminate $\hat{\mathcal{S}}$ in (53). We obtain

$$\begin{aligned} \hat{\mathcal{U}}^{[s]} &= \mathcal{X} \times_1 \left(\hat{\mathcal{U}}_1^{[s]} \cdot \hat{\mathcal{U}}_1^{[s]H} \right) \dots \times_R \left(\hat{\mathcal{U}}_R^{[s]} \cdot \hat{\mathcal{U}}_R^{[s]H} \right) \\ &\quad \times_{R+1} \left(\hat{\Sigma}_s^{-1} \cdot \hat{\mathcal{U}}_{R+1}^{[s]H} \right) \\ &= \mathcal{X} \times_1 \hat{\mathcal{T}}_1 \dots \times_R \hat{\mathcal{T}}_R \times_{R+1} \left(\hat{\Sigma}_s^{-1} \cdot \hat{\mathcal{U}}_{R+1}^{[s]H} \right), \end{aligned} \quad (56)$$

where we have introduced the short hand notation $\hat{\mathcal{T}}_r = \hat{\mathcal{U}}_r^{[s]} \cdot \hat{\mathcal{U}}_r^{[s]H}$. The next step is to compute the matrix $[\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T$. Inserting (57) and using (5), we obtain

$$[\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T = \left(\hat{\mathcal{T}}_1 \otimes \dots \otimes \hat{\mathcal{T}}_R \right) \cdot [\mathcal{X}]_{(R+1)}^T \cdot \hat{\mathcal{U}}_{R+1}^{[s]*} \cdot \hat{\Sigma}_s^{-1}. \quad (58)$$

As pointed out in Section III-A, the link between the measurement matrix \mathbf{X} and the measurement tensor \mathcal{X} is given by $\mathbf{X} = [\mathcal{X}]_{(R+1)}^T$. Therefore, their SVDs (cf. (18) and (23)) are linked through the following identities

$$\hat{\mathcal{U}}_s = \hat{\mathcal{V}}_{R+1}^{[s]*}, \quad \hat{\mathcal{U}}_n = \hat{\mathcal{V}}_{R+1}^{[n]*}, \quad \hat{\mathcal{V}}_s = \hat{\mathcal{U}}_{R+1}^{[s]*}, \quad \hat{\mathcal{V}}_n = \hat{\mathcal{U}}_{R+1}^{[n]*}.$$

Consequently we can write

$$\begin{aligned} [\mathcal{X}]_{(R+1)}^T \cdot \hat{\mathcal{U}}_{R+1}^{[s]*} \cdot \hat{\Sigma}_s^{-1} &= \mathbf{X} \cdot \hat{\mathcal{V}}_s \cdot \hat{\Sigma}_s^{-1} \\ &= \hat{\mathcal{U}} \cdot \hat{\Sigma} \cdot \hat{\mathcal{V}}^H \cdot \hat{\mathcal{V}}_s \cdot \hat{\Sigma}_s^{-1} = \hat{\mathcal{U}}_s \cdot \hat{\Sigma}_s \cdot \hat{\Sigma}_s^{-1} = \hat{\mathcal{U}}_s. \end{aligned} \quad (59)$$

Finally, inserting (59) into (58) yields

$$[\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T = \left(\hat{\mathcal{T}}_1 \otimes \dots \otimes \hat{\mathcal{T}}_R \right) \cdot \hat{\mathcal{U}}_s, \quad (60)$$

which is the desired result. □

Corollary 1. *A corollary which follows from this theorem is that the exact subspace \mathcal{U}_s satisfies the following identity*

$$\mathcal{U}_s = (\mathcal{T}_1 \otimes \dots \otimes \mathcal{T}_R) \cdot \mathcal{U}_s. \quad (61)$$

Proof: The corollary follows by considering the special case where $\mathcal{X} = \mathcal{X}_0$ and hence $\hat{\mathcal{T}}_r = \mathcal{T}_r$ as well as $\hat{\mathcal{U}}_s = \mathcal{U}_s$. For this case we also have $[\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T = [\mathcal{U}^{[s]}]_{(R+1)}^T = \mathcal{U}_s$, where the last identity resembles the fact that in the noise-free case, the HOSVD-based subspace estimate coincides with the SVD-based subspace estimate. ■

APPENDIX B
PROOF OF THEOREM 2

We start by inserting $\hat{\mathcal{U}}_s = \mathcal{U}_s + \Delta\mathcal{U}_s$ and $\hat{\mathcal{T}}_r = \mathcal{T}_r + \Delta\mathcal{T}_r$ into (25). Then we obtain

$$\begin{aligned} [\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T &= [(\mathcal{T}_1 + \Delta\mathcal{T}_1) \otimes \dots \otimes (\mathcal{T}_R + \Delta\mathcal{T}_R)] \cdot (\mathcal{U}_s + \Delta\mathcal{U}_s) \\ &= \underbrace{[\mathcal{T}_1 \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s}_{\mathcal{U}_s} + [\mathcal{T}_1 \otimes \dots \otimes \mathcal{T}_R] \cdot \Delta\mathcal{U}_s \\ &\quad + [\Delta\mathcal{T}_1 \otimes \mathcal{T}_2 \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s + \dots \\ &\quad + [\mathcal{T}_1 \otimes \mathcal{T}_2 \otimes \dots \otimes \Delta\mathcal{T}_R] \cdot \mathcal{U}_s + \mathcal{O}\{\Delta^2\}, \end{aligned} \quad (62)$$

since all terms that contain more than one perturbation term can be absorbed into $\mathcal{O}\{\Delta^2\}$. The first term in (62) represents the exact signal subspace (cf. Corollary 1), hence the remaining terms are the first order expansion of $[\Delta\hat{\mathcal{U}}^{[s]}]_{(R+1)}^T$. As the first term of this expansion already agrees with Theorem 2, we still need to show that for the remaining terms we have for $r = 1, 2, \dots, R$

$$\begin{aligned} [\mathcal{T}_1 \otimes \dots \otimes \Delta\mathcal{T}_r \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s &= \\ [\mathcal{T}_1 \otimes \dots \otimes (\mathcal{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} \cdot \mathcal{U}_r^{[s]H}) \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s + \mathcal{O}\{\Delta^2\}. \end{aligned} \quad (63)$$

As a first step, we expand the left-hand side of (63) by applying Corollary 1

$$\begin{aligned} &[\mathcal{T}_1 \otimes \dots \otimes \Delta\mathcal{T}_r \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s \\ &= [\mathcal{T}_1 \otimes \dots \otimes \Delta\mathcal{T}_r \otimes \dots \otimes \mathcal{T}_R] \cdot [\mathcal{T}_1 \otimes \dots \otimes \mathcal{T}_r \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s \\ &= [(\mathcal{T}_1 \cdot \mathcal{T}_1) \otimes \dots \otimes (\Delta\mathcal{T}_r \cdot \mathcal{T}_r) \otimes \dots \otimes (\mathcal{T}_R \cdot \mathcal{T}_R)] \cdot \mathcal{U}_s \\ &= [\mathcal{T}_1 \otimes \dots \otimes (\Delta\mathcal{T}_r \cdot \mathcal{T}_r) \otimes \dots \otimes \mathcal{T}_R] \cdot \mathcal{U}_s, \end{aligned} \quad (64)$$

where we have used the fact that the matrices \mathcal{T}_r are projection matrices and hence idempotent, i.e., $\mathcal{T}_r \cdot \mathcal{T}_r = \mathcal{T}_r$. What remains to be shown is that $\Delta\mathcal{T}_r \cdot \mathcal{T}_r = \mathcal{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} \cdot \mathcal{U}_r^{[s]H} + \mathcal{O}\{\Delta^2\}$. Since $\hat{\mathcal{T}}_r = \hat{\mathcal{U}}_r^{[s]} \cdot \hat{\mathcal{U}}_r^{[s]H}$ and $\hat{\mathcal{U}}_r^{[s]} = \mathcal{U}_r^{[s]} + \Delta\mathcal{U}_r^{[s]}$, a first order expansion for $\Delta\mathcal{T}_r$ is obtained via

$$\begin{aligned} \hat{\mathcal{T}}_r &= \left(\mathcal{U}_r^{[s]} + \Delta\mathcal{U}_r^{[s]} \right) \cdot \left(\mathcal{U}_r^{[s]H} + \Delta\mathcal{U}_r^{[s]H} \right) \\ &= \mathcal{T}_r + \mathcal{U}_r^{[s]} \cdot \Delta\mathcal{U}_r^{[s]H} + \Delta\mathcal{U}_r^{[s]} \cdot \mathcal{U}_r^{[s]H} + \mathcal{O}\{\Delta^2\} \\ &\Rightarrow \Delta\mathcal{T}_r = \mathcal{U}_r^{[s]} \cdot \Delta\mathcal{U}_r^{[s]H} + \Delta\mathcal{U}_r^{[s]} \cdot \mathcal{U}_r^{[s]H} + \mathcal{O}\{\Delta^2\}, \end{aligned} \quad (65)$$

where in general we have $\Delta\mathcal{U}_r^{[s]} = \mathcal{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} + \mathcal{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]} + \mathcal{O}\{\Delta^2\}$ (cf. (24)). Using this expansion in (65) we obtain

$$\Delta\mathcal{T}_r \cdot \mathcal{T}_r$$

$$\begin{aligned}
&= \mathbf{U}_r^{[s]} \cdot \left(\mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} + \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]} \right)^H \cdot \mathbf{T}_r \\
&\quad + \left(\mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} + \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]} \right) \cdot \mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r + \mathcal{O}\{\Delta^2\} \\
&= \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[n]H} \cdot \mathbf{U}_r^{[n]H} \cdot \mathbf{T}_r + \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]H} \cdot \mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r \\
&\quad + \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]} \cdot \mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r + \mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} \cdot \mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r + \mathcal{O}\{\Delta^2\} \\
&= \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[n]H} \cdot \underbrace{\mathbf{U}_r^{[n]H} \cdot \mathbf{T}_r}_{\mathbf{0}_{M_r - p_r \times M_r}} + \underbrace{\mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r}_{\mathbf{0}_{p_r \times p_r}} \cdot \mathbf{U}_r^{[s]} \cdot \mathbf{\Gamma}_r^{[s]H} \cdot \mathbf{T}_r \\
&\quad + \mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} \cdot \underbrace{\mathbf{U}_r^{[s]H} \cdot \mathbf{T}_r}_{\mathbf{U}_r^{[s]H}} + \mathcal{O}\{\Delta^2\} \\
&= \mathbf{U}_r^{[n]} \cdot \mathbf{\Gamma}_r^{[n]} \cdot \mathbf{U}_r^{[s]H} + \mathcal{O}\{\Delta^2\}, \tag{66}
\end{aligned}$$

which is the desired result. Note that $\mathbf{\Gamma}_r^{[s]} + \mathbf{\Gamma}_r^{[s]H} = \mathbf{0}_{p_r \times p_r}$ follows from the fact that $\mathbf{\Gamma}_r^{[s]}$ is a skew-Hermitian matrix (which is apparent from its definition shown in (24)). This completes the proof of the theorem. \square

APPENDIX C PROOF OF THEOREM 3

For R -D Standard ESPRIT, the explicit first-order expansion of the estimation error for the k -th spatial frequency in the r -th mode in terms of the signal subspace estimation error $\Delta \mathbf{U}_s$ is given by (30). This error can be expressed in terms of the perturbation (noise) matrix \mathbf{N} by inserting (19). We obtain

$$\begin{aligned}
\Delta \mu_k^{(r)} &= \text{Im} \left\{ \mathbf{r}_k^{(r)T} \cdot \text{vec} \{ \Delta \mathbf{U}_s \} \right\} + \mathcal{O}\{\Delta^2\} \\
&= \text{Im} \left\{ \mathbf{r}_k^{(r)T} \cdot \mathbf{W}_{\text{mat}} \cdot \text{vec} \{ \mathbf{N} \} \right\} + \mathcal{O}\{\Delta^2\} \tag{67} \\
\mathbf{r}_k^{(r)T} &= \mathbf{q}_k^{(r)T} \otimes \left(\mathbf{p}_k^{(r)T} \cdot \left(\tilde{\mathbf{J}}_1^{(r)} \cdot \mathbf{U}_s \right)^+ \cdot \left[\tilde{\mathbf{J}}_2^{(r)} / \lambda_k^{(r)} - \tilde{\mathbf{J}}_1^{(r)} \right] \right)
\end{aligned}$$

$$\mathbf{W}_{\text{mat}} = \left(\mathbf{\Sigma}_s^{-1} \cdot \mathbf{V}_s^T \right) \otimes \left(\mathbf{U}_n \cdot \mathbf{U}_n^H \right),$$

which follows directly by applying property (1) to (30) and to (19). In order to expand $\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\}$ using (67), we observe that for arbitrary complex vectors $\mathbf{z}_1, \mathbf{z}_2$ we have $\text{Im} \{ \mathbf{z}_1^T \cdot \mathbf{z}_2 \} = \text{Im} \{ \mathbf{z}_1 \}^T \cdot \text{Re} \{ \mathbf{z}_2 \} + \text{Re} \{ \mathbf{z}_1 \}^T \cdot \text{Im} \{ \mathbf{z}_2 \}$ and hence

$$\begin{aligned}
\text{Im} \{ \mathbf{z}_1^T \cdot \mathbf{z}_2 \}^2 &= \text{Im} \{ \mathbf{z}_1 \}^T \cdot \text{Re} \{ \mathbf{z}_2 \} \cdot \text{Re} \{ \mathbf{z}_2 \}^T \cdot \text{Im} \{ \mathbf{z}_1 \} \\
&\quad + \text{Re} \{ \mathbf{z}_1 \}^T \cdot \text{Im} \{ \mathbf{z}_2 \} \cdot \text{Im} \{ \mathbf{z}_2 \}^T \cdot \text{Re} \{ \mathbf{z}_1 \} \\
&\quad + \text{Im} \{ \mathbf{z}_1 \}^T \cdot \text{Re} \{ \mathbf{z}_2 \} \cdot \text{Im} \{ \mathbf{z}_2 \}^T \cdot \text{Re} \{ \mathbf{z}_1 \} \\
&\quad + \text{Re} \{ \mathbf{z}_1 \}^T \cdot \text{Im} \{ \mathbf{z}_2 \} \cdot \text{Re} \{ \mathbf{z}_2 \}^T \cdot \text{Im} \{ \mathbf{z}_1 \} \tag{68}
\end{aligned}$$

Using (67) in $\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\}$ and applying (C) for $\mathbf{z}_1^T = \mathbf{r}_k^{(r)T} \cdot \mathbf{W}_{\text{mat}}$ and $\mathbf{z}_2 = \text{vec} \{ \mathbf{N} \} = \mathbf{n}$ we find

$$\begin{aligned}
\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\} &= \mathbb{E} \left\{ \text{Im} \{ \mathbf{z}_1^T \} \cdot \text{Re} \{ \mathbf{n} \} \cdot \text{Re} \{ \mathbf{n} \}^T \cdot \text{Im} \{ \mathbf{z}_1 \} \right\} \\
&\quad + \mathbb{E} \left\{ \text{Re} \{ \mathbf{z}_1^T \} \cdot \text{Im} \{ \mathbf{n} \} \cdot \text{Im} \{ \mathbf{n} \}^T \cdot \text{Re} \{ \mathbf{z}_1 \} \right\} \\
&\quad + \mathbb{E} \left\{ \text{Im} \{ \mathbf{z}_1^T \} \cdot \text{Re} \{ \mathbf{n} \} \cdot \text{Im} \{ \mathbf{n} \}^T \cdot \text{Re} \{ \mathbf{z}_1 \} \right\} \\
&\quad + \mathbb{E} \left\{ \text{Re} \{ \mathbf{z}_1^T \} \cdot \text{Im} \{ \mathbf{n} \} \cdot \text{Re} \{ \mathbf{n} \}^T \cdot \text{Im} \{ \mathbf{z}_1 \} \right\} \tag{69}
\end{aligned}$$

Since the only random quantity in (69) is the vector of noise samples \mathbf{n} , we can move \mathbf{z}_1 out of the expectation operator. We are then left with the covariance matrices of the real part and the imaginary part of the noise, respectively, as well as with the cross-covariance matrix between the real and the imaginary part. To proceed we require the following lemma:

Lemma 2. *Let \mathbf{n} be a zero mean random vector with covariance matrix $\mathbf{R}_{\text{nn}} = \mathbb{E} \{ \mathbf{n} \cdot \mathbf{n}^H \}$ and pseudo-covariance matrix $\mathbf{C}_{\text{nn}} = \mathbb{E} \{ \mathbf{n} \cdot \mathbf{n}^T \}$. Then, the covariance matrices of the real part of \mathbf{n} , the imaginary part of \mathbf{n} and the cross-covariance between the real and the imaginary part of \mathbf{n} are given by*

$$\begin{aligned}
\mathbf{R}_{\text{nn}}^{(\text{R},\text{R})} &\doteq \mathbb{E} \left\{ \text{Re} \{ \mathbf{n} \} \cdot \text{Re} \{ \mathbf{n} \}^T \right\} = \frac{1}{2} \text{Re} \{ \mathbf{R}_{\text{nn}} + \mathbf{C}_{\text{nn}} \} \\
\mathbf{R}_{\text{nn}}^{(\text{I},\text{I})} &\doteq \mathbb{E} \left\{ \text{Im} \{ \mathbf{n} \} \cdot \text{Im} \{ \mathbf{n} \}^T \right\} = \frac{1}{2} \text{Re} \{ \mathbf{R}_{\text{nn}} - \mathbf{C}_{\text{nn}} \} \\
\mathbf{R}_{\text{nn}}^{(\text{R},\text{I})} &\doteq \mathbb{E} \left\{ \text{Re} \{ \mathbf{n} \} \cdot \text{Im} \{ \mathbf{n} \}^T \right\} = -\frac{1}{2} \text{Im} \{ \mathbf{R}_{\text{nn}} - \mathbf{C}_{\text{nn}} \} \\
\mathbf{R}_{\text{nn}}^{(\text{I},\text{R})} &\doteq \mathbb{E} \left\{ \text{Im} \{ \mathbf{n} \} \cdot \text{Re} \{ \mathbf{n} \}^T \right\} = \frac{1}{2} \text{Im} \{ \mathbf{R}_{\text{nn}} + \mathbf{C}_{\text{nn}} \}.
\end{aligned}$$

Proof: To prove this Lemma we expand \mathbf{R}_{nn} and \mathbf{C}_{nn} by inserting $\mathbf{n} = \text{Re} \{ \mathbf{n} \} + j \text{Im} \{ \mathbf{n} \}$. We then obtain

$$\mathbf{R}_{\text{nn}} = \mathbf{R}_{\text{nn}}^{(\text{R},\text{R})} + \mathbf{R}_{\text{nn}}^{(\text{I},\text{I})} + j \left(\mathbf{R}_{\text{nn}}^{(\text{I},\text{R})} - \mathbf{R}_{\text{nn}}^{(\text{R},\text{I})} \right) \tag{70}$$

$$\mathbf{C}_{\text{nn}} = \mathbf{R}_{\text{nn}}^{(\text{R},\text{R})} - \mathbf{R}_{\text{nn}}^{(\text{I},\text{I})} + j \left(\mathbf{R}_{\text{nn}}^{(\text{I},\text{R})} + \mathbf{R}_{\text{nn}}^{(\text{R},\text{I})} \right). \tag{71}$$

Since $\mathbf{R}_{\text{nn}}^{(\text{R},\text{R})}$, $\mathbf{R}_{\text{nn}}^{(\text{R},\text{I})}$, $\mathbf{R}_{\text{nn}}^{(\text{I},\text{R})}$, and $\mathbf{R}_{\text{nn}}^{(\text{I},\text{I})}$ are real-valued, the solution of (70) and (71) is straightforward. \blacksquare

Using Lemma 2 in (69) we obtain for the mean square error

$$\begin{aligned}
\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\} &= \frac{1}{2} \left(\text{Im} \{ \mathbf{z}_1^T \} \cdot \text{Re} \{ \mathbf{R}_{\text{nn}} + \mathbf{C}_{\text{nn}} \} \cdot \text{Im} \{ \mathbf{z}_1 \} \right. \\
&\quad + \text{Re} \{ \mathbf{z}_1^T \} \cdot \text{Re} \{ \mathbf{R}_{\text{nn}} - \mathbf{C}_{\text{nn}} \} \cdot \text{Re} \{ \mathbf{z}_1 \} \\
&\quad + \text{Im} \{ \mathbf{z}_1^T \} \cdot \text{Im} \{ -\mathbf{R}_{\text{nn}} + \mathbf{C}_{\text{nn}} \} \cdot \text{Re} \{ \mathbf{z}_1 \} \\
&\quad \left. + \text{Re} \{ \mathbf{z}_1^T \} \cdot \text{Im} \{ \mathbf{R}_{\text{nn}} + \mathbf{C}_{\text{nn}} \} \cdot \text{Im} \{ \mathbf{z}_1 \} \right) \tag{72}
\end{aligned}$$

Finally, (72) can be expressed in more compact form as

$$\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\} = \frac{1}{2} \left(\mathbf{z}_1^H \cdot \mathbf{R}_{\text{nn}}^T \cdot \mathbf{z}_1 - \text{Re} \left\{ \mathbf{z}_1^T \cdot \mathbf{C}_{\text{nn}}^T \cdot \mathbf{z}_1 \right\} \right) \tag{73}$$

for $\mathbf{z}_1 = \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_k^{(r)}$, which is the desired result. \square

Note that Gaussianity is not needed for these properties to hold. Consequently, the MSE expressions are still valid if the noise is not Gaussian. We only need it to be zero mean. Also, note that for the special case of circularly symmetric white noise we have $\mathbf{R}_{\text{nn}} = \sigma_n^2 \cdot \mathbf{I}_{MN}$ and $\mathbf{C}_{\text{nn}} = \mathbf{0}_{MN \times MN}$ and hence the MSE simplifies into

$$\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\} = \frac{\sigma_n^2}{2} \cdot \|\mathbf{z}_1\|_2^2 = \frac{\sigma_n^2}{2} \cdot \left\| \mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_k^{(r)} \right\|_2^2. \tag{74}$$

The procedure for 2-D Standard Tensor-ESPRIT is in fact quite similar. The first step is to express the estimation

error in $\mu_k^{(r)}$ in terms of the perturbation $\mathbf{n} = \text{vec}\{\mathbf{N}\} = \text{vec}\{[\mathcal{N}]_{(3)}^T\}$ (cf. (13)). This expression takes the form

$$\begin{aligned}\Delta\mu_k^{(r)} &= \text{Im}\left\{\mathbf{r}_k^{(r)T} \cdot \text{vec}\left\{\left[\Delta\hat{\mathbf{U}}^{[s]}_{(R+1)}\right]^T\right\}\right\} + \mathcal{O}\{\Delta^2\} \\ &= \text{Im}\left\{\mathbf{r}_k^{(r)T} \cdot \mathbf{W}_{\text{ten}} \cdot \text{vec}\{\mathbf{N}\}\right\} + \mathcal{O}\{\Delta^2\}\end{aligned}\quad (75)$$

since $\left[\Delta\hat{\mathbf{U}}^{[s]}_{(R+1)}\right]^T$ depends linearly on $\text{vec}\{\mathbf{N}\}$. Due to the fact that (75) has the same form as (67), the second step to expand the MSE expressions follows the same lines as for R-D Standard ESPRIT, which immediately shows that the MSE becomes

$$\mathbb{E}\left\{(\Delta\mu_k^{(r)})^2\right\} = \frac{1}{2}\left(\mathbf{z}_1^H \cdot \mathbf{R}_{\text{nn}}^T \cdot \mathbf{z}_1 - \text{Re}\left\{\mathbf{z}_1^T \cdot \mathbf{C}_{\text{nn}}^T \cdot \mathbf{z}_1\right\}\right) \quad (76)$$

for $\mathbf{z}_1 = \mathbf{W}_{\text{ten}}^T \cdot \mathbf{r}_k^{(r)}$. Therefore, the final step is finding an explicit expression for \mathbf{W}_{ten} which satisfies

$$\left[\Delta\hat{\mathbf{U}}^{[s]}_{(3)}\right]^T = \mathbf{W}_{\text{ten}} \cdot \text{vec}\{\mathbf{N}\} + \mathcal{O}\{\Delta^2\}. \quad (77)$$

Recall from Theorem 2 that for $R = 2$, the HOSVD-based signal subspace estimation error $\left[\Delta\hat{\mathbf{U}}^{[s]}_{(R+1)}\right]$ can be expanded into

$$\begin{aligned}\left[\Delta\hat{\mathbf{U}}^{[s]}_{(3)}\right]^T &= (\mathbf{T}_1 \otimes \mathbf{T}_2) \cdot \Delta\mathbf{U}_s + \left(\left[\Delta\mathbf{U}_1^{[s]} \cdot \mathbf{U}_1^{[s]H}\right] \otimes \mathbf{T}_2\right) \cdot \mathbf{U}_s \\ &\quad + \left(\mathbf{T}_1 \otimes \left[\Delta\mathbf{U}_2^{[s]} \cdot \mathbf{U}_2^{[s]H}\right]\right) \cdot \mathbf{U}_s + \mathcal{O}\{\Delta^2\},\end{aligned}\quad (78)$$

where $\Delta\mathbf{U}_s$, $\Delta\mathbf{U}_1^{[s]}$, and $\Delta\mathbf{U}_2^{[s]}$ are given by

$$\begin{aligned}\Delta\mathbf{U}_s &= \mathbf{U}_n \cdot \mathbf{U}_n^H \cdot \mathbf{N} \cdot \mathbf{V}_s \cdot \Sigma_s^{-1} \\ &= \mathbf{V}_3^{[n]*} \cdot \mathbf{V}_3^{[n]T} \cdot \mathbf{N} \cdot \mathbf{U}_3^{[s]*} \cdot \Sigma_3^{[s]-1} \quad \text{and} \\ \Delta\mathbf{U}_r^{[s]} &= \mathbf{U}_r^{[n]} \cdot \mathbf{U}_r^{[n]H} \cdot [\mathcal{N}]_{(r)} \cdot \mathbf{V}_r^{[s]} \cdot \Sigma_r^{[s]-1} \quad \text{for } r = 1, 2.\end{aligned}\quad (79)$$

The first term in (78) is easily vectorized by applying property (1) which yields the first term of \mathbf{W}_{ten} as

$$\begin{aligned}\text{vec}\{(\mathbf{T}_1 \otimes \mathbf{T}_2) \cdot \Delta\mathbf{U}_s\} &= \text{vec}\left\{(\mathbf{T}_1 \otimes \mathbf{T}_2) \cdot \mathbf{V}_3^{[n]*} \cdot \mathbf{V}_3^{[n]T} \cdot \mathbf{N} \cdot \mathbf{U}_3^{[s]*} \cdot \Sigma_3^{[s]-1}\right\} \\ &= \left(\mathbf{U}_3^{[s]*} \cdot \Sigma_3^{[s]-1}\right)^T \otimes \left[(\mathbf{T}_1 \otimes \mathbf{T}_2) \cdot \mathbf{V}_3^{[n]*} \cdot \mathbf{V}_3^{[n]T}\right] \cdot \text{vec}\{\mathbf{N}\} \\ &= \left(\Sigma_3^{[s]-1} \cdot \mathbf{U}_3^{[s]H}\right) \otimes \left[(\mathbf{T}_1 \otimes \mathbf{T}_2) \cdot \mathbf{V}_3^{[n]*} \cdot \mathbf{V}_3^{[n]T}\right] \cdot \text{vec}\{\mathbf{N}\}.\end{aligned}$$

However, for the second term in (78) we get

$$\begin{aligned}\text{vec}\left\{\left(\left[\mathbf{U}_1^{[n]} \cdot \mathbf{U}_1^{[n]H} \cdot [\mathcal{N}]_{(1)} \cdot \mathbf{V}_1^{[s]} \cdot \Sigma_1^{[s]-1} \cdot \mathbf{U}_1^{[s]H}\right] \otimes \mathbf{T}_2\right) \cdot \mathbf{U}_s\right\} \\ = \left(\mathbf{U}_s^T \otimes \mathbf{I}_M\right) \cdot \text{vec}\left\{\left[\mathbf{U}_1^{[n]} \cdot \mathbf{U}_1^{[n]H} \cdot [\mathcal{N}]_{(1)} \cdot \mathbf{V}_1^{[s]} \cdot \Sigma_1^{[s]-1} \cdot \mathbf{U}_1^{[s]H}\right] \otimes \mathbf{T}_2\right\}\end{aligned}$$

by inserting (79) for $\Delta\mathbf{U}_1^{[s]}$. To proceed we need to rewrite the vectorization of a Kronecker product. After straightforward calculations we obtain

$$\left(\mathbf{U}_s^T \otimes \mathbf{I}_M\right) \cdot$$

$$\begin{aligned}&\text{vec}\left\{\left[\mathbf{U}_1^{[n]} \cdot \mathbf{U}_1^{[n]H} \cdot [\mathcal{N}]_{(1)} \cdot \mathbf{V}_1^{[s]} \cdot \Sigma_1^{[s]-1} \cdot \mathbf{U}_1^{[s]H}\right] \otimes \mathbf{T}_2\right\} \\ &= \left(\mathbf{U}_s^T \otimes \mathbf{I}_M\right) \cdot \bar{\mathbf{T}}_2 \cdot \text{vec}\left\{\left[\mathbf{U}_1^{[n]} \cdot \mathbf{U}_1^{[n]H} \cdot [\mathcal{N}]_{(1)} \cdot \mathbf{V}_1^{[s]} \cdot \Sigma_1^{[s]-1} \cdot \mathbf{U}_1^{[s]H}\right]\right\} \quad (80) \\ &= \left(\mathbf{U}_s^T \otimes \mathbf{I}_M\right) \cdot \bar{\mathbf{T}}_2 \cdot \left[\left(\mathbf{V}_1^{[s]} \cdot \Sigma_1^{[s]-1} \cdot \mathbf{U}_1^{[s]H}\right)^T \otimes \left(\mathbf{U}_1^{[n]} \cdot \mathbf{U}_1^{[n]H}\right)\right] \text{vec}\{[\mathcal{N}]_{(1)}\}\end{aligned}$$

where the matrix $\bar{\mathbf{T}}_2$ is constructed from the columns of \mathbf{T}_2 given by $\mathbf{t}_{2,m}$ for $m = 1, 2, \dots, M_2$ in the following manner

$$\bar{\mathbf{T}}_2 = \mathbf{I}_{M_1} \otimes \begin{bmatrix} \mathbf{I}_{M_1} \otimes \mathbf{t}_{2,1} \\ \vdots \\ \mathbf{I}_{M_1} \otimes \mathbf{t}_{2,M_2} \end{bmatrix}. \quad (81)$$

The final step is to rearrange the elements of $\text{vec}\{[\mathcal{N}]_{(1)}\}$ so that they appear in the same order as in $\text{vec}\{\mathbf{N}\}$. However, since $\mathbf{N} = [\mathcal{N}]_{(3)}^T$, this can easily be achieved in the following manner

$$\text{vec}\{[\mathcal{N}]_{(1)}\} = \mathbf{K}_{M_2 \times (M_1 \cdot N)} \cdot \text{vec}\{\mathbf{N}\} \quad (82)$$

where $\mathbf{K}_{M_2 \times (M_1 \cdot N)}$ is the commutation matrix (cf. equation (7)). This completes the derivation of the second term of \mathbf{W}_{ten} . The third term is obtained in a similar manner. In this case, no permutation is needed, since $\text{vec}\{[\mathcal{N}]_{(2)}\} = \text{vec}\{[\mathcal{N}]_{(3)}^T\} = \text{vec}\{\mathbf{N}\}$. \square

APPENDIX D PROOF OF THEOREM 4

As pointed out in Section V-D, the inclusion of Forward-Backward-Averaging leads to a very similar model, where all quantities originating from the noise-free observation \mathbf{X}_0 (or \mathcal{X}_0) are replaced by the corresponding quantities for $\mathbf{X}_0^{(\text{fba})}$ (or $\mathcal{X}_0^{(\text{fba})}$). Since for Theorem 3 it was only assumed that the desired signal component is superimposed by a zero mean noise contribution, it is directly applicable.

The only point we need to derive are the covariance matrix and the pseudo-covariance matrix of the forward-backward averaged noise $\mathbf{n}^{(\text{fba})} \doteq \text{vec}\{\mathbf{N}^{(\text{fba})}\}$, which are needed for the MSE expressions. To this end, we can express $\mathbf{n}^{(\text{fba})}$ as

$$\begin{aligned}\text{vec}\{\mathbf{N}^{(\text{fba})}\} &= \text{vec}\{[\mathbf{N}, \quad \mathbf{\Pi}_M \cdot \mathbf{N}^* \cdot \mathbf{\Pi}_N]\} \\ &= \begin{bmatrix} \text{vec}\{\mathbf{N}\} \\ (\mathbf{\Pi}_N \otimes \mathbf{\Pi}_M) \cdot \text{vec}\{\mathbf{N}^*\} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{n} \\ \mathbf{\Pi}_{NM} \cdot \mathbf{n}^* \end{bmatrix}\end{aligned}\quad (83)$$

Equation (83) allows us to express the covariance matrix and the pseudo-covariance matrix of $\mathbf{n}^{(\text{fba})}$ via the covariance matrix and the pseudo-covariance matrix of \mathbf{n} . We obtain

$$\begin{aligned}\mathbb{E}\left\{\mathbf{n}^{(\text{fba})} \cdot \mathbf{n}^{(\text{fba})H}\right\} &= \begin{bmatrix} \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^H\} & \mathbb{E}\{\mathbf{n} \cdot \mathbf{n}^T\} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbb{E}\{\mathbf{n}^* \cdot \mathbf{n}^H\} & \mathbf{\Pi}_{MN} \cdot \mathbb{E}\{\mathbf{n}^* \cdot \mathbf{n}^T\} \cdot \mathbf{\Pi}_{MN} \end{bmatrix}\end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} \mathbf{R}_{nn} & \mathbf{C}_{nn} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbf{C}_{nn}^* & \mathbf{\Pi}_{MN} \cdot \mathbf{R}_{nn}^* \cdot \mathbf{\Pi}_{MN} \end{bmatrix} \\
&\mathbb{E} \left\{ \mathbf{n}^{(\text{fba})} \cdot \mathbf{n}^{(\text{fba})\text{T}} \right\} \\
&= \begin{bmatrix} \mathbb{E} \{ \mathbf{n} \cdot \mathbf{n}^{\text{T}} \} & \mathbb{E} \{ \mathbf{n} \cdot \mathbf{n}^{\text{H}} \} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbb{E} \{ \mathbf{n}^* \cdot \mathbf{n}^{\text{T}} \} & \mathbf{\Pi}_{MN} \cdot \mathbb{E} \{ \mathbf{n}^* \cdot \mathbf{n}^{\text{H}} \} \cdot \mathbf{\Pi}_{MN} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{C}_{nn} & \mathbf{R}_{nn} \cdot \mathbf{\Pi}_{MN} \\ \mathbf{\Pi}_{MN} \cdot \mathbf{R}_{nn}^* & \mathbf{\Pi}_{MN} \cdot \mathbf{C}_{nn}^* \cdot \mathbf{\Pi}_{MN} \end{bmatrix}
\end{aligned}$$

This completes the proof of the theorem. \square

APPENDIX E PROOF OF THEOREM 5

Without regularization, the cost function for 1-D Structured Least Squares can be expressed as [8]

$$\hat{\Psi}_{\text{SLS}} = \arg \min_{\Psi, \Delta \bar{\mathbf{U}}_s} \left\| \mathbf{J}_1 \cdot (\hat{\mathbf{U}}_s + \Delta \bar{\mathbf{U}}_s) \cdot \Psi - \mathbf{J}_2 \cdot (\hat{\mathbf{U}}_s + \Delta \bar{\mathbf{U}}_s) \right\|_{\text{F}}^2. \quad \mathbf{F}_{\text{SLS}} = \left[\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \mathbf{U}_s), (\Psi^{\text{T}} \otimes \mathbf{J}_1) - (\mathbf{I}_d \otimes \mathbf{J}_2) \right] \quad (84)$$

where we have used $\Delta \bar{\mathbf{U}}_s$ only to avoid confusion with the $\Delta \mathbf{U}_s$ associated to the estimation error in $\hat{\mathbf{U}}_s$. Note that the cost function is solved in an iterative manner starting with $\Delta \bar{\mathbf{U}}_s = \mathbf{0}_{M \times d}$ and with $\Psi = \Psi_{\text{LS}}$, where $\Psi_{\text{LS}} = (\mathbf{J}_1 \cdot \hat{\mathbf{U}}_s)^+ \cdot (\mathbf{J}_2 \cdot \hat{\mathbf{U}}_s)$ represents the LS solution to the shift invariance equation. As we compute only a single iteration we find one update term for $\hat{\mathbf{U}}_s$ and one for Ψ_{LS} which we denote as $\Delta \mathbf{U}_{\text{s,SLS}}$ and $\Delta \Psi_{\text{SLS}}$ (i.e., $\Delta \mathbf{U}_{\text{s,SLS}}$ represents the $\Delta \bar{\mathbf{U}}_s$ which minimizes the linearized version of (84)). In other words, the cost function becomes

$$\begin{aligned}
\hat{\Psi}_{\text{SLS}} &= \hat{\Psi}_{\text{LS}} + \Delta \Psi_{\text{SLS}} \quad \text{where} \quad (85) \\
\Delta \Psi_{\text{SLS}} &= \arg \min_{\Delta \Psi, \Delta \bar{\mathbf{U}}_s} \left\| \mathbf{J}_1 \cdot (\hat{\mathbf{U}}_s + \Delta \bar{\mathbf{U}}_s) \cdot (\Psi_{\text{LS}} + \Delta \Psi) \right. \\
&\quad \left. - \mathbf{J}_2 \cdot (\hat{\mathbf{U}}_s + \Delta \bar{\mathbf{U}}_s) \right\|_{\text{F}}^2 \\
&= \arg \min_{\Delta \Psi, \Delta \bar{\mathbf{U}}_s} \left\| \mathbf{R}_{\text{LS}} + \mathbf{J}_1 \cdot \Delta \bar{\mathbf{U}}_s \cdot \Psi_{\text{LS}} + \mathbf{J}_1 \cdot \hat{\mathbf{U}}_s \cdot \Delta \Psi \right. \\
&\quad \left. - \mathbf{J}_2 \cdot \Delta \bar{\mathbf{U}}_s + \mathcal{O} \{ \Delta^2 \} \right\|_{\text{F}}^2 \quad (86)
\end{aligned}$$

where we have defined the matrix $\mathbf{R}_{\text{LS}} = \mathbf{J}_1 \cdot \hat{\mathbf{U}}_s \cdot \Psi_{\text{LS}} - \mathbf{J}_2 \cdot \hat{\mathbf{U}}_s$ which contains the residual error in the shift invariance equation after the LS fit. Since (86) is linearized by skipping the quadratic terms in $\mathcal{O} \{ \Delta^2 \}$, it is easily solved by an LS fit. To express the result in closed-form we vectorize (86) using the fact that $\|\mathbf{A}\|_{\text{F}} = \|\text{vec} \{ \mathbf{A} \} \|_2$ and obtain

$$\begin{aligned}
\Delta \psi_{\text{SLS}} &= \arg \min_{\text{vec} \{ \Delta \Psi \}, \text{vec} \{ \Delta \bar{\mathbf{U}}_s \}} \left\| \mathbf{r}_{\text{LS}} + \left(\hat{\Psi}_{\text{LS}}^{\text{T}} \otimes \mathbf{J}_1 \right) \cdot \text{vec} \{ \Delta \bar{\mathbf{U}}_s \} \right. \\
&\quad \left. + \left(\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \hat{\mathbf{U}}_s) \right) \cdot \text{vec} \{ \Delta \Psi \} \right. \\
&\quad \left. - (\mathbf{I}_d \otimes \mathbf{J}_2) \cdot \text{vec} \{ \Delta \bar{\mathbf{U}}_s \} + \mathcal{O} \{ \Delta^2 \} \right\|_2^2 \quad (87)
\end{aligned}$$

$$\Rightarrow \Delta \psi_{\text{SLS}} = \arg \min_{\text{vec} \{ \Delta \Psi \}, \text{vec} \{ \Delta \bar{\mathbf{U}}_s \}} \left\| \mathbf{r}_{\text{LS}} + \hat{\mathbf{F}}_{\text{SLS}} \cdot \begin{bmatrix} \text{vec} \{ \Delta \Psi \} \\ \text{vec} \{ \Delta \bar{\mathbf{U}}_s \} \end{bmatrix} + \mathcal{O} \{ \Delta^2 \} \right\|_2^2 \quad (88)$$

$$\Rightarrow \begin{bmatrix} \Delta \psi_{\text{SLS}} \\ \Delta \mathbf{u}_{\text{s,SLS}} \end{bmatrix} = -\hat{\mathbf{F}}_{\text{SLS}}^+ \cdot \mathbf{r}_{\text{LS}} \quad (89)$$

where we have introduced the vectorized quantities $\mathbf{r}_{\text{LS}} = \text{vec} \{ \mathbf{R}_{\text{LS}} \}$, $\Delta \psi_{\text{SLS}} = \text{vec} \{ \Delta \Psi_{\text{SLS}} \}$, and $\Delta \mathbf{u}_{\text{s,SLS}} = \text{vec} \{ \Delta \mathbf{U}_{\text{s,SLS}} \}$, respectively. We have also skipped the quadratic terms $\mathcal{O} \{ \Delta^2 \}$ in (88) for the solution in (89). Moreover, the matrix $\hat{\mathbf{F}}_{\text{SLS}} \in \mathbb{C}^{(M-1)d \times (d^2 + M \cdot d)}$ becomes

$$\hat{\mathbf{F}}_{\text{SLS}} = \left[\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \hat{\mathbf{U}}_s), \left(\hat{\Psi}_{\text{LS}}^{\text{T}} \otimes \mathbf{J}_1 \right) - (\mathbf{I}_d \otimes \mathbf{J}_2) \right]. \quad (90)$$

Therefore, our next goal is to find a first order expansion of $\Delta \psi_{\text{SLS}}$ in (89). This looks difficult at first sight as it involves an expansion of a pseudo-inverse due to $\hat{\mathbf{F}}_{\text{SLS}}$. However, this step simplifies significantly by realizing that $\hat{\mathbf{F}}_{\text{SLS}}$ can be expressed as $\hat{\mathbf{F}}_{\text{SLS}} = \mathbf{F}_{\text{SLS}} + \Delta \hat{\mathbf{F}}_{\text{SLS}}$, where

$$\begin{aligned}
\mathbf{F}_{\text{SLS}} &= \left[\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \mathbf{U}_s), (\Psi^{\text{T}} \otimes \mathbf{J}_1) - (\mathbf{I}_d \otimes \mathbf{J}_2) \right] \quad (91) \\
\Delta \mathbf{F}_{\text{SLS}} &= \left[\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \Delta \mathbf{U}_s), \left(\Delta \Psi_{\text{LS}}^{\text{T}} \otimes \mathbf{J}_1 \right) \right]
\end{aligned}$$

where $\hat{\Psi}_{\text{LS}} = \Psi + \Delta \Psi_{\text{LS}}$. Since \mathbf{F}_{SLS} is not random (i.e., only dependent on \mathbf{X}_0 but not on \mathbf{N}) and $\Delta \Psi_{\text{LS}} = \mathbf{0}_{d \times d} + \mathcal{O} \{ \Delta \}$, i.e., at least linear in the perturbation, we have

$$\hat{\mathbf{F}}_{\text{SLS}}^+ = \mathbf{F}_{\text{SLS}}^+ + \mathcal{O} \{ \Delta \}. \quad (92)$$

This relation only describes the “zero-th” term of the expansion of $\hat{\mathbf{F}}_{\text{SLS}}^+$. However, as we see below, the linear term is not needed for a first order expansion of $\Delta \psi_{\text{SLS}}$. Continuing with (89), the second term of the right-hand side is given by \mathbf{r}_{LS} for which we can write

$$\begin{aligned}
\mathbf{r}_{\text{LS}} &= \text{vec} \left\{ \mathbf{J}_1 \cdot \hat{\mathbf{U}}_s \cdot \hat{\Psi}_{\text{LS}} - \mathbf{J}_2 \cdot \hat{\mathbf{U}}_s \right\} \\
&= \text{vec} \left\{ \mathbf{J}_1 \cdot (\mathbf{U}_s + \Delta \mathbf{U}_s) \cdot (\Psi + \Delta \Psi_{\text{LS}}) \right. \\
&\quad \left. - \mathbf{J}_2 \cdot (\mathbf{U}_s + \Delta \mathbf{U}_s) \right\} \\
&= \text{vec} \left\{ \mathbf{J}_1 \cdot \Delta \mathbf{U}_s \cdot \Psi + \mathbf{J}_1 \cdot \mathbf{U}_s \cdot \Delta \Psi_{\text{LS}} \right. \\
&\quad \left. - \mathbf{J}_2 \cdot \Delta \mathbf{U}_s + \mathcal{O} \{ \Delta^2 \} \right\} \quad (93)
\end{aligned}$$

Moreover, as shown in [17], $\Delta \Psi_{\text{LS}}$ can be expressed in terms of $\Delta \mathbf{U}_s$ via

$$\begin{aligned}
\Delta \Psi_{\text{LS}} &= (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_2 \cdot \Delta \mathbf{U}_s \\
&\quad - (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_1 \cdot \Delta \mathbf{U}_s \cdot \Psi + \mathcal{O} \{ \Delta^2 \}. \quad (94)
\end{aligned}$$

Using this expansion in (93) we obtain

$$\begin{aligned}
\mathbf{r}_{\text{LS}} &= \text{vec} \left\{ \mathbf{J}_1 \cdot \Delta \mathbf{U}_s \cdot \Psi \right. \\
&\quad \left. + \text{vec} \left\{ \mathbf{J}_1 \cdot \mathbf{U}_s \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_2 \cdot \Delta \mathbf{U}_s \right\} \right. \\
&\quad \left. - \text{vec} \left\{ \mathbf{J}_1 \cdot \mathbf{U}_s \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_1 \cdot \Delta \mathbf{U}_s \cdot \Psi \right\} \right. \\
&\quad \left. - \text{vec} \left\{ \mathbf{J}_2 \cdot \Delta \mathbf{U}_s \right\} + \mathcal{O} \{ \Delta^2 \} \right\} \\
&= \mathbf{W}_{\text{R,U}} \cdot \text{vec} \{ \Delta \mathbf{U}_s \} + \mathcal{O} \{ \Delta^2 \} \quad \text{where} \quad (95) \\
\mathbf{W}_{\text{R,U}} &= \left(\Psi^{\text{T}} \otimes \mathbf{J}_1 \right) + \mathbf{I}_d \otimes \left(\mathbf{J}_1 \cdot \mathbf{U}_s (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_2 \right) \\
&\quad - \Psi^{\text{T}} \otimes \left(\mathbf{J}_1 \cdot \mathbf{U}_s (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \mathbf{J}_1 \right) - (\mathbf{I}_d \otimes \mathbf{J}_2).
\end{aligned}$$

Using (95) and (92) in (89) we find that

$$\begin{bmatrix} \Delta \psi_{\text{SLS}} \\ \Delta \mathbf{u}_{\text{s,SLS}} \end{bmatrix} = -\hat{\mathbf{F}}_{\text{SLS}}^+ \cdot \mathbf{r}_{\text{LS}}$$

$$\begin{aligned}
&= -(\mathbf{F}_{\text{SLS}}^+ + \mathcal{O}\{\Delta\}) \cdot (\mathbf{W}_{\text{R,U}} \cdot \text{vec}\{\Delta\mathbf{U}_s\} + \mathcal{O}\{\Delta^2\}) \\
&= -\mathbf{F}_{\text{SLS}}^+ \cdot \mathbf{W}_{\text{R,U}} \cdot \text{vec}\{\Delta\mathbf{U}_s\} + \mathcal{O}\{\Delta^2\}. \quad (96)
\end{aligned}$$

Note that from (91) it follows that \mathbf{F}_{SLS} has full row-rank and hence its pseudo-inverse can be expressed as $\mathbf{F}_{\text{SLS}}^+ = \mathbf{F}_{\text{SLS}}^H \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1}$. This allows us to extract $\Delta\psi_{\text{SLS}}$ from (96) since $\Delta\mathbf{u}_{\text{s,SLS}}$ is not explicitly needed as long as only one SLS iteration is performed. We obtain

$$\begin{aligned}
\Delta\psi_{\text{SLS}} &= -\left(\mathbf{I}_d \otimes (\mathbf{J}_1 \cdot \mathbf{U}_s)^H\right) \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1} \\
&\quad \cdot \mathbf{W}_{\text{R,U}} \cdot \text{vec}\{\Delta\mathbf{U}_s\} + \mathcal{O}\{\Delta^2\} \quad (97)
\end{aligned}$$

The final step in SLS-based ESPRIT is to replace the LS-based estimate $\hat{\Psi}_{\text{LS}}$ by $\hat{\Psi}_{\text{SLS}} = \hat{\Psi}_{\text{LS}} + \Delta\Psi_{\text{SLS}} = \Psi + \Delta\Psi_{\text{LS}} + \Delta\Psi_{\text{SLS}}$. Following the first-order expansion for Standard ESPRIT from [17] we obtain

$$\begin{aligned}
\Delta\mu_k &= \text{Im}\left\{\mathbf{p}_k^T \cdot (\Delta\Psi_{\text{LS}} + \Delta\Psi_{\text{SLS}}) \cdot \mathbf{q}_k\right\} / e^{j\mu_k} + \mathcal{O}\{\Delta^2\} \\
&= \text{Im}\left\{\mathbf{p}_k^T \cdot \Delta\Psi_{\text{LS}} \cdot \mathbf{q}_k\right\} / e^{j\mu_k} \\
&\quad + \text{Im}\left\{(\mathbf{q}_k^T \otimes \mathbf{p}_k^T) \cdot \Delta\psi_{\text{SLS}}\right\} / e^{j\mu_k} + \mathcal{O}\{\Delta^2\} \quad (98)
\end{aligned}$$

Since the first term is exactly the same as for LS-based 1-D Standard ESPRIT, we can use the result from [17]. Inserting the expansion for $\Delta\psi_{\text{SLS}}$ from (97) we obtain

$$\begin{aligned}
\Delta\mu_k &= \text{Im}\left\{\mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \left(\frac{\mathbf{J}_2}{e^{j\mu_k}} - \mathbf{J}_1\right) \cdot \Delta\mathbf{U}_s \cdot \mathbf{q}_k\right\} \\
&\quad - \text{Im}\left\{(\mathbf{q}_k^T \otimes \mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^H) \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1} \right. \\
&\quad \left. \cdot \mathbf{W}_{\text{R,U}} \cdot \text{vec}\{\Delta\mathbf{U}_s\}\right\} / e^{j\mu_k} + \mathcal{O}\{\Delta^2\} \quad (99)
\end{aligned}$$

Finally, rearranging the first term as $\mathbf{a}^T \cdot \mathbf{B} \cdot \mathbf{c} = (\mathbf{c}^T \otimes \mathbf{a}^T) \cdot \text{vec}\{\mathbf{B}\}$ we have

$$\begin{aligned}
\Delta\mu_k &= \text{Im}\left\{\mathbf{q}_k^T \otimes \left(\mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \left(\frac{\mathbf{J}_2}{e^{j\mu_k}} - \mathbf{J}_1\right)\right) \right. \\
&\quad \left. \cdot \text{vec}\{\Delta\mathbf{U}_s\}\right\} \\
&\quad - \text{Im}\left\{(\mathbf{q}_k^T \otimes \mathbf{p}_k^T \cdot \frac{(\mathbf{J}_1 \cdot \mathbf{U}_s)^H}{e^{j\mu_k}}) \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1} \right. \\
&\quad \left. \cdot \mathbf{W}_{\text{R,U}} \cdot \text{vec}\{\Delta\mathbf{U}_s\}\right\} + \mathcal{O}\{\Delta^2\} \\
&= \text{Im}\left\{\mathbf{r}_{k,\text{SLS}}^T \cdot \text{vec}\{\Delta\mathbf{U}_s\}\right\} + \mathcal{O}\{\Delta^2\}, \quad \text{where} \\
\mathbf{r}_{k,\text{SLS}}^T &= \mathbf{q}_k^T \otimes \left[\mathbf{p}_k^T \cdot (\mathbf{J}_1 \cdot \mathbf{U}_s)^+ \cdot \left(\frac{\mathbf{J}_2}{e^{j\mu_k}} - \mathbf{J}_1\right)\right] \\
&\quad - \left(\mathbf{q}_k^T \otimes \left[\mathbf{p}_k^T \cdot \frac{(\mathbf{J}_1 \cdot \mathbf{U}_s)^H}{e^{j\mu_k}}\right]\right) \cdot (\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H)^{-1} \cdot \mathbf{W}_{\text{R,U}}, \quad (100)
\end{aligned}$$

which is the desired result (42). Equation (43) follows from (42) by inserting the first order expansion for $\text{vec}\{\Delta\mathbf{U}_s\}$ in terms of $\text{vec}\{\mathbf{N}\}$ as shown in Appendix C. \square

APPENDIX F PROOF OF THEOREM 6

This theorem consists of several parts which we address in separate subsections.

A. MSE for Standard ESPRIT

We start by simplifying the MSE expression for 1-D Standard ESPRIT. In the case of a single source we can write

$$\mathbf{X}_0 = \mathbf{a}(\mu) \cdot \mathbf{s}^T, \quad (101)$$

where $\mathbf{a} \in \mathbb{C}^{M \times 1}$ is the array steering vector and $\mathbf{s} \in \mathbb{C}^{N \times 1}$ contains the source symbols. Let $\hat{P}_T = \|\mathbf{s}\|_2^2 / N$ be the empirical source power. Furthermore, since we assume a ULA of isotropic elements, $\mathbf{a}(\mu)$ is given by $\mathbf{a}(\mu) = [1, e^{j\mu}, e^{2j\mu}, \dots, e^{(M-1)j\mu}]$. Note that $\|\mathbf{a}(\mu)\|_2^2 = M$. For notational convenience, we drop the explicit dependence of \mathbf{a} on μ and write just $\mathbf{a}(\mu) = \mathbf{a}$ in the sequel. The selection matrices \mathbf{J}_1 and \mathbf{J}_2 are then chosen as

$$\mathbf{J}_1 = [\mathbf{I}_{M-1} \quad \mathbf{0}_{M-1 \times 1}] \quad \mathbf{J}_2 = [\mathbf{0}_{M-1 \times 1} \quad \mathbf{I}_{M-1}] \quad (102)$$

for maximum overlap, i.e., $M^{(\text{sel})} = M - 1$. Since (101) is a rank-one matrix, we can directly relate the subspaces to the array steering vector and the source symbol matrix, namely

$$\mathbf{U}_s = \mathbf{u}_s = \frac{\mathbf{a}}{\|\mathbf{a}\|_2} = \frac{1}{\sqrt{M}} \cdot \mathbf{a} \quad (103)$$

$$\mathbf{V}_s = \mathbf{v}_s = \frac{\mathbf{s}^*}{\|\mathbf{s}\|_2} = \frac{1}{\sqrt{\hat{P}_T \cdot N}} \cdot \mathbf{s}^* \quad (104)$$

$$\Sigma_s = \sigma_s = \sqrt{M \cdot N \cdot \hat{P}_T}. \quad (105)$$

For the MSE expression from Theorem 3 we also require the quantity $\mathbf{U}_n \cdot \mathbf{U}_n^H$, which resembles a projection matrix on the noise subspace. However, since the signal subspace is spanned by \mathbf{a} we can write $\mathbf{U}_n \cdot \mathbf{U}_n^H = \mathbf{Pr}_a^\perp = \mathbf{I}_M - \frac{\mathbf{a} \cdot \mathbf{a}^H}{\|\mathbf{a}\|_2^2} = \mathbf{I}_M - \frac{1}{M} \cdot \mathbf{a} \cdot \mathbf{a}^H$. The MSE expression for 1-D Standard ESPRIT also include the eigenvectors of Ψ which for the special case discussed here is scalar and given by $\Psi = e^{j\mu}$. Consequently, we have $\mathbf{p}_k = \mathbf{q}_k = 1$ for the eigenvectors.

Combining these expressions and inserting into the special case of (33) for white noise, which is shown in equation (74), we have $\mathbb{E}\{(\Delta\mu_k^{(r)})^2\} = \sigma_n^2/2 \cdot \|\mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}\|_2^2 = \sigma_n^2/2 \cdot \|\mathbf{r}^T \cdot \mathbf{W}_{\text{mat}}\|_2^2$ with

$$\mathbf{r}_k^{(r)} = \mathbf{r} = \left[\left(\mathbf{J}_1 \frac{\mathbf{a}}{\sqrt{M}} \right)^+ (\mathbf{J}_2 / e^{j\mu} - \mathbf{J}_1) \right]^T \quad (106)$$

$$\mathbf{W}_{\text{mat}} = \left(\frac{1}{\sqrt{M \cdot N \cdot \hat{P}_T}} \cdot \frac{\mathbf{s}^H}{\sqrt{\hat{P}_T \cdot N}} \right) \otimes \mathbf{Pr}_a^\perp \quad (107)$$

Note that \mathbf{W}_{mat} is the Kronecker product of a $1 \times N$ vector and an $M \times M$ matrix. Hence, $\mathbf{r}^T \cdot \mathbf{W}_{\text{mat}}$ can be written as $\mathbf{r}^T \cdot \mathbf{W}_{\text{mat}} = \tilde{\mathbf{s}}^T \otimes \tilde{\mathbf{a}}^T$, where

$$\tilde{\mathbf{s}}^T = \frac{1}{\sqrt{M \cdot N \cdot \hat{P}_T}} \cdot \frac{\mathbf{s}^H}{\sqrt{\hat{P}_T \cdot N}} \quad (108)$$

$$\tilde{\mathbf{a}}^T = \left(\mathbf{J}_1 \frac{\mathbf{a}}{\sqrt{M}} \right)^+ \left(\frac{\mathbf{J}_2}{e^{j\mu}} - \mathbf{J}_1 \right) \cdot \mathbf{Pr}_a^\perp \quad (109)$$

Therefore, the MSE can be expressed as $\mathbb{E}\{(\Delta\mu_k^{(r)})^2\} = \sigma_n^2/2 \cdot \|\tilde{\mathbf{s}}^T \otimes \tilde{\mathbf{a}}^T\|_2^2$, which is equal to $\mathbb{E}\{(\Delta\mu_k^{(r)})^2\} = \sigma_n^2/2 \cdot \|\tilde{\mathbf{s}}^T\|_2^2 \cdot \|\tilde{\mathbf{a}}^T\|_2^2$.

Since $\tilde{\mathbf{s}}^T$ is a scaled version of \mathbf{s}^H and $\|\mathbf{s}^H\|_2^2 = N \cdot \hat{P}_T$ we find that the first term in the MSE expression can conveniently be expressed as

$$\|\tilde{\mathbf{s}}^T\|_2^2 = \frac{1}{M \cdot N \cdot \hat{P}_T} \cdot \frac{\hat{P}_T \cdot N}{\hat{P}_T \cdot N} = \frac{1}{M \cdot N \cdot \hat{P}_T}. \quad (110)$$

Next, we proceed to simplify $\tilde{\mathbf{a}}^T$ further. To this end, we expand the pseudo-inverse of $\mathbf{J}_1 \cdot \mathbf{a}$ using the rule $\mathbf{x}^+ = \mathbf{x}^H / \|\mathbf{x}\|_2^2$ and multiply the brackets out. After straightforward algebraic manipulations we obtain

$$\begin{aligned} \tilde{\mathbf{a}}^T &= \frac{\sqrt{M}}{M-1} (\tilde{\mathbf{a}}_1^T - \tilde{\mathbf{a}}_2^T), \quad \text{where} \\ \tilde{\mathbf{a}}_1^T &= \mathbf{a}^H \cdot \mathbf{J}_1^H \cdot \mathbf{J}_2 / e^{j\mu} \quad \text{and} \quad \tilde{\mathbf{a}}_2^T = \mathbf{a}^H \cdot \mathbf{J}_1^H \cdot \mathbf{J}_1. \end{aligned} \quad (111)$$

Since we have $\mathbf{a}^H = [1, e^{-j\mu}, e^{-2j\mu}, \dots, e^{-(M-1)j\mu}]$ it is easy to see that

$$\begin{aligned} \tilde{\mathbf{a}}_1^T &= [0, e^{-j\mu}, e^{-2j\mu}, \dots, e^{-(M-2)j\mu}, e^{-(M-1)j\mu}] \\ \tilde{\mathbf{a}}_2^T &= [1, e^{-j\mu}, e^{-2j\mu}, \dots, e^{-(M-2)j\mu}, 0] \quad \text{and hence} \\ \tilde{\mathbf{a}}_1^T - \tilde{\mathbf{a}}_2^T &= [-1, 0, \dots, 0, e^{-(M-1)j\mu}]. \end{aligned} \quad (112)$$

Consequently, we find $\|\tilde{\mathbf{a}}^T\|_2^2 = \frac{M}{(M-1)^2} \cdot 2$. Combining this result with (110) we finally have

$$\mathbb{E} \left\{ (\Delta \mu_k^{(r)})^2 \right\} = \frac{\sigma_n^2}{2} \cdot \|\tilde{\mathbf{s}}^T\|_2^2 \cdot \|\tilde{\mathbf{a}}^T\|_2^2 \quad (113)$$

$$\begin{aligned} &= \frac{\sigma_n^2}{2} \cdot \frac{1}{M \cdot N \cdot \hat{P}_T} \cdot 2 \cdot \frac{M}{(M-1)^2} \\ &= \frac{\sigma_n^2}{N \cdot \hat{P}_T} \cdot \frac{1}{(M-1)^2} \end{aligned} \quad (114)$$

which is the desired result. \square

B. MSE for Unitary ESPRIT

The second part of the theorem is to show that for a single source, the MSE for Unitary ESPRIT is the same as the MSE for Standard ESPRIT. Firstly, we expand $\mathbf{X}_0^{(\text{fba})}$ and find

$$\begin{aligned} \mathbf{X}_0^{(\text{fba})} &= [\mathbf{a} \cdot \mathbf{s}^T, \quad \mathbf{\Pi}_M \cdot \mathbf{a}^* \cdot \mathbf{s}^H \cdot \mathbf{\Pi}_N] \\ &= \mathbf{a} \cdot [\mathbf{s}^T \quad e^{-j\mu(M-1)} \cdot \mathbf{s}^H \cdot \mathbf{\Pi}_N] = \mathbf{a} \cdot \bar{\mathbf{s}}^T \end{aligned} \quad (115)$$

where we have used the fact that for our ULA we have $\mathbf{\Pi}_M \cdot \mathbf{a}^* = \mathbf{a} \cdot e^{-j\mu(M-1)}$ and we have defined $\bar{\mathbf{s}}$ to be

$$\bar{\mathbf{s}} = \begin{bmatrix} \mathbf{s} \\ e^{-j\mu(M-1)} \cdot \mathbf{\Pi}_N \cdot \mathbf{s}^* \end{bmatrix} \quad (116)$$

Note that $\bar{\mathbf{s}}^H \bar{\mathbf{s}} = \mathbf{s}^H \cdot \mathbf{s} + \mathbf{s}^T \cdot \mathbf{\Pi}_N \cdot \mathbf{\Pi}_N \cdot \mathbf{s}^* = 2 \cdot \mathbf{s}^H \cdot \mathbf{s}$. As for Standard ESPRIT, we relate (115) to its SVD and obtain

$$\begin{aligned} \mathbf{u}_s^{(\text{fba})} &= \frac{\mathbf{a}}{\sqrt{M}} = \mathbf{u}_s, \quad \mathbf{v}_s^{(\text{fba})} = \frac{\bar{\mathbf{s}}^*}{\sqrt{2 \cdot N \cdot \hat{P}_T}}, \\ \sigma_s^{(\text{fba})} &= \sqrt{2 \cdot M \cdot N \cdot \hat{P}_T}. \end{aligned} \quad (117)$$

An important consequence we can draw from (117) is that the column space \mathbf{u}_s remains unaffected from the forward-backward-averaging. Therefore we also have $\mathbf{U}_n^{(\text{fba})} = \mathbf{U}_n$ and hence $\mathbf{U}_n^{(\text{fba})} \mathbf{U}_n^{(\text{fba})H} = \mathbf{I}_M - \frac{\mathbf{a} \cdot \mathbf{a}^H}{M}$. However, the equivalence of Standard ESPRIT and Unitary ESPRIT is still not obvious

since Forward-Backward Averaging destroys the circular symmetry of the noise, which leads to an additional term in the MSE expressions. Following the lines of the derivation for 1-D Standard ESPRIT we can show that

$$\mathbf{r}^{(\text{fba})T} \cdot \mathbf{W}_{\text{mat}}^{(\text{fba})} = \tilde{\mathbf{s}}^T \otimes \tilde{\mathbf{a}}^T, \quad (118)$$

where $\tilde{\mathbf{s}}$ is given by

$$\tilde{\mathbf{s}} = \frac{1}{\sqrt{2 \cdot M \cdot N \cdot \hat{P}_T}} \cdot \frac{\bar{\mathbf{s}}^*}{\sqrt{2 \cdot N \cdot \hat{P}_T}} \quad (119)$$

and $\tilde{\mathbf{a}}$ is the same as in the derivation for 1-D Standard ESPRIT (cf. equation (109)).

According to Theorem 4, the MSE for Unitary ESPRIT can be computed as

$$\frac{\sigma_n^2}{2} \cdot \left(\mathbf{z}^T \cdot \mathbf{z}^* - \text{Re} \{ \mathbf{z}^T \cdot \mathbf{\Pi}_{2MN} \cdot \mathbf{z} \} \right), \quad (120)$$

for $\mathbf{z}^T = \mathbf{r}^{(\text{fba})T} \cdot \mathbf{W}_{\text{mat}}^{(\text{fba})}$, where we have inserted $\mathbf{R}_{\text{nn}}^{(\text{fba})} = \sigma_n^2 \cdot \mathbf{I}_{2MN}$ and $\mathbf{C}_{\text{nn}}^{(\text{fba})} = \sigma_n^2 \cdot \mathbf{\Pi}_{2MN}$ since we are considering the special case of circularly symmetric white noise. Using (118) and the fact that $\mathbf{\Pi}_{2MN} = \mathbf{\Pi}_{2N} \otimes \mathbf{\Pi}_M$, this expression can be written into

$$\frac{\sigma_n^2}{2} \cdot \left(\|\tilde{\mathbf{s}}\|_2^2 \|\tilde{\mathbf{a}}\|_2^2 - \tilde{\mathbf{s}}^T \cdot \mathbf{\Pi}_{2N} \cdot \tilde{\mathbf{s}}^* \cdot \tilde{\mathbf{a}}^T \cdot \mathbf{\Pi}_M \cdot \tilde{\mathbf{a}} \right). \quad (121)$$

Since $\tilde{\mathbf{a}}$ is the same as in (111) we know that $\|\tilde{\mathbf{a}}\|_2^2 = \frac{2M}{(M-1)^2}$. Moreover, $\|\tilde{\mathbf{s}}\|_2^2 = \frac{1}{2 \cdot M \cdot N \cdot \hat{P}_T}$ follows directly from (119). For the second term in (121) we have

$$\begin{aligned} \tilde{\mathbf{s}}^T \cdot \mathbf{\Pi}_{2N} \cdot \tilde{\mathbf{s}}^* &= \frac{1}{2MN\hat{P}_T} \cdot \frac{1}{2N\hat{P}_T} \cdot \bar{\mathbf{s}}^H \cdot \mathbf{\Pi}_{2N} \cdot \bar{\mathbf{s}} \\ &= \frac{1}{2MN\hat{P}_T} \cdot \frac{1}{2N\hat{P}_T} \cdot \left[\mathbf{s}^H \cdot \mathbf{s} \cdot e^{j\mu(M-1)} + \mathbf{s}^T \cdot \mathbf{s}^* \cdot e^{j\mu(M-1)} \right] \\ &= \frac{1}{2MN\hat{P}_T} \cdot e^{j\mu(M-1)} \end{aligned} \quad (122)$$

Similarly we can simplify $\tilde{\mathbf{a}}^T \cdot \mathbf{\Pi}_M \cdot \tilde{\mathbf{a}}$ by using (111) and (112). We obtain

$$\tilde{\mathbf{a}}^T \cdot \mathbf{\Pi}_M \cdot \tilde{\mathbf{a}} = -\frac{2M}{(M-1)^2} \cdot e^{-(M-1)j\mu}. \quad (123)$$

Combining the results from (122) and (123) into (121) we finally obtain for the MSE

$$\begin{aligned} &\frac{\sigma_n^2}{2} \cdot \left(\frac{1}{2MN\hat{P}_T} \cdot \frac{2M}{(M-1)^2} \right. \\ &\quad \left. + \frac{1}{2MN\hat{P}_T} \cdot e^{j\mu(M-1)} \cdot \frac{2M}{(M-1)^2} \cdot e^{-(M-1)j\mu} \right) \\ &= \frac{\sigma_n^2}{2} \cdot \left(\frac{1}{N\hat{P}_T} \cdot \frac{1}{(M-1)^2} + \frac{1}{N\hat{P}_T} \cdot \frac{1}{(M-1)^2} \right) \end{aligned} \quad (124)$$

$$= \frac{\sigma_n^2}{N \cdot \hat{P}_T} \cdot \frac{1}{(M-1)^2}, \quad (125)$$

which is equal to the result for 1-D Standard ESPRIT from (114) and hence proves this part of the theorem. \square

C. Cramér-Rao Bound

The third part of the theorem is to simplify the deterministic Cramér-Rao Bound (CRB) for the special case of a single source. To this end, a closed-form expression for the deterministic CRB for this setting is given by [34]

$$C = \frac{\sigma_n^2}{2 \cdot N} \cdot \text{Re} \left\{ \left[\mathbf{D}^H \cdot \mathbf{Pr}_A^\perp \cdot \mathbf{D} \right] \odot \hat{\mathbf{R}}_S^T \right\}^{-1} \quad (126)$$

where $\hat{\mathbf{R}}_S = \frac{1}{N} \cdot \mathbf{S} \cdot \mathbf{S}^H$ is the sample covariance matrix of the source symbols, $\mathbf{D} \in \mathbb{C}^{M \times d}$ is the matrix of partial derivatives of the array steering vectors with respect to the parameters of interest, and $\mathbf{Pr}_A^\perp = \mathbf{I}_M - \mathbf{A} \cdot (\mathbf{A}^H \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^H$. In the case $d = 1$, we have $\hat{\mathbf{R}}_S = \|\mathbf{s}\|_2^2 / N = \hat{P}_T$ and the CRB expression simplifies into

$$C = \frac{\sigma_n^2}{2 \cdot N \cdot \hat{P}_T} \cdot \text{Re} \left\{ \mathbf{d}^H \cdot \left(\mathbf{I}_M - \frac{\mathbf{a} \cdot \mathbf{a}^H}{M} \right) \cdot \mathbf{d} \right\}^{-1} \quad (127)$$

$$= \frac{1}{2 \cdot \hat{\rho}} \cdot \text{Re} \left\{ \mathbf{d}^H \cdot \mathbf{d} - \frac{1}{M} \cdot \mathbf{d}^H \cdot \mathbf{a} \cdot \mathbf{a}^H \cdot \mathbf{d} \right\}^{-1} \quad (128)$$

$$= \frac{1}{2 \cdot \hat{\rho}} \cdot \left[\mathbf{d}^H \cdot \mathbf{d} - \frac{1}{M} \cdot |\mathbf{d}^H \cdot \mathbf{a}|^2 \right]^{-1} \quad (129)$$

Since for a ULA the array steering vector can be expressed as $\mathbf{a} = [1 \quad e^{j\mu} \quad e^{2j\mu} \quad \dots \quad e^{(M-1)j\mu}]$ we have

$$\mathbf{d} = \frac{\partial \mathbf{a}}{\partial \mu} = j \cdot [0 \quad e^{j\mu} \quad 2 \cdot e^{2j\mu} \quad \dots \quad (M-1) \cdot e^{(M-1)j\mu}]. \quad (130)$$

Consequently the terms $\mathbf{d}^H \cdot \mathbf{d}$ and $\mathbf{d}^H \cdot \mathbf{a}$ become

$$\mathbf{d}^H \cdot \mathbf{d} = \sum_{m=0}^{M-1} m^2 = \frac{1}{6} \cdot (M-1) \cdot M \cdot (2M-1) \quad (131)$$

$$\mathbf{d}^H \cdot \mathbf{a} = -j \sum_{m=0}^{M-1} m = -j \cdot \frac{1}{2} \cdot (M-1) \cdot M \quad (132)$$

Using these expressions in (129), we obtain

$$C = \frac{1}{2 \cdot \hat{\rho}} \cdot \left[\frac{1}{6} \cdot (M-1) \cdot M \cdot (2M-1) - \frac{1}{M} \cdot \left| -j \cdot \frac{1}{2} \cdot (M-1) \cdot M \right|^2 \right]^{-1} \quad (133)$$

$$= \frac{1}{2 \cdot \hat{\rho}} \cdot \left[\frac{1}{12} \cdot (M-1) \cdot M \cdot (M+1) \right]^{-1} \quad (134)$$

$$= \frac{1}{\hat{\rho}} \cdot \frac{6}{(M-1) \cdot M \cdot (M+1)},$$

which is the desired result. \square

APPENDIX G PROOF OF THEOREM 7

As shown in Theorem 5, the MSE for SLS in the special case of circularly symmetric white noise can be expressed as

$$\mathbb{E} \left\{ (\Delta \mu_{k,\text{SLS}})^2 \right\} = \frac{\sigma_n^2}{2} \cdot \|\mathbf{W}_{\text{mat}} \cdot \mathbf{r}_{k,\text{SLS}}\|_2^2, \quad (135)$$

where $\mathbf{r}_{k,\text{SLS}} = \mathbf{r}_{k,\text{LS}} - \Delta \mathbf{r}_{k,\text{SLS}}$,

$$\Delta \mathbf{r}_{k,\text{SLS}}^T = \left(\mathbf{q}_k^T \otimes \left[\mathbf{p}_k^T \cdot \frac{(\mathbf{J}_1 \cdot \mathbf{U}_s)^H}{e^{j\mu_k}} \right] \right)$$

$$\cdot \left(\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H \right)^{-1} \cdot \mathbf{W}_{\text{R,U}},$$

and $\mathbf{W}_{\text{R,U}}$ as well as \mathbf{F}_{SLS} are defined in Theorem 5. For a single source, we have $\mathbf{p}_k = \mathbf{q}_k = 1$, $\Psi = \Psi = e^{j\mu}$, and $\mathbf{U}_s = \mathbf{a}/\sqrt{M}$, and therefore $\Delta \mathbf{r}_{k,\text{SLS}}^T = \Delta \mathbf{r}_{\text{SLS}}^T$ simplifies to

$$\Delta \mathbf{r}_{\text{SLS}}^T = \frac{(\mathbf{J}_1 \cdot \mathbf{a})^H}{\sqrt{M} \cdot e^{j\mu}} \cdot \left(\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H \right)^{-1} \cdot \mathbf{W}_{\text{R,U}} \quad (136)$$

$$\begin{aligned} \mathbf{W}_{\text{R,U}} &= (e^{j\mu} \cdot \mathbf{J}_1) + \left(\mathbf{J}_1 \cdot \mathbf{a} (\mathbf{J}_1 \cdot \mathbf{a})^+ \cdot \mathbf{J}_2 \right) \\ &\quad - e^{j\mu} \cdot \left(\mathbf{J}_1 \cdot \mathbf{a} (\mathbf{J}_1 \cdot \mathbf{a})^+ \cdot \mathbf{J}_1 \right) - \mathbf{J}_2 \\ \mathbf{F}_{\text{SLS}} &= \left[\mathbf{J}_1 \cdot \frac{\mathbf{a}}{\sqrt{M}}, e^{j\mu} \cdot \mathbf{J}_1 - \mathbf{J}_2 \right] \end{aligned}$$

We can write $\mathbf{W}_{\text{R,U}}$ as

$$\mathbf{W}_{\text{R,U}} = \left(\mathbf{J}_1 \cdot \mathbf{a} (\mathbf{J}_1 \cdot \mathbf{a})^+ - \mathbf{I}_{M-1} \right) \cdot (\mathbf{J}_2 - e^{j\mu} \cdot \mathbf{J}_1) \quad (137)$$

Moreover, we need to simplify the term $\left(\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H \right)^{-1}$. It is easily verified that $\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H$ can be written as

$$\begin{aligned} \mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H &= \text{diag}(\mathbf{J}_1 \cdot \mathbf{a}) \cdot \mathbf{G} \cdot \text{diag}(\mathbf{J}_1 \cdot \mathbf{a})^H \quad (138) \\ \mathbf{G} &= \frac{1}{M} \cdot \mathbf{1}_{(M-1) \times (M-1)} + 2 \cdot \mathbf{I}_{M-1} - \mathbf{J}_1 \cdot \mathbf{J}_2^H - \mathbf{J}_2 \cdot \mathbf{J}_1^H \quad (139) \end{aligned}$$

Equation (138) shows that the inverse of $\mathbf{F}_{\text{SLS}} \cdot \mathbf{F}_{\text{SLS}}^H$ can be expressed as $\text{diag}(\mathbf{J}_1 \cdot \mathbf{a}) \cdot \mathbf{G}^{-1} \cdot \text{diag}(\mathbf{J}_1 \cdot \mathbf{a})^H$. To proceed further, we require the following Lemma:

Lemma 3. *The inverse of the matrix \mathbf{G} defined in (139) is given by the following expression*

$$[\mathbf{G}^{-1}]_{(m_1, m_2)} = \begin{cases} \frac{1}{M} \cdot \left((M - m_1) \cdot m_2 - 3 \cdot \frac{m_1 \cdot (M - m_1) \cdot m_2 \cdot (M - m_2)}{M^2 + 11} \right) & m_1 \geq m_2 \\ \frac{1}{M} \cdot \left(m_1 \cdot (M - m_2) - 3 \cdot \frac{m_1 \cdot (M - m_1) \cdot m_2 \cdot (M - m_2)}{M^2 + 11} \right) & m_1 < m_2 \end{cases} \quad (140)$$

$$m_1, m_2 = 1, 2, \dots, M-1.$$

Proof: To prove this Lemma it is sufficient to multiply \mathbf{G}^{-1} in (140) with \mathbf{G} defined in (139) and show that the result is an identity matrix. \blacksquare

Collecting our intermediate results from (136), (138), and (137) we have for $\Delta \mathbf{r}_{\text{SLS}}^T$

$$\begin{aligned} \Delta \mathbf{r}_{\text{SLS}}^T &= \frac{(\mathbf{J}_1 \mathbf{a})^H}{\sqrt{M} \cdot e^{j\mu}} \cdot \text{diag}(\mathbf{J}_1 \cdot \mathbf{a}) \cdot \mathbf{G}^{-1} \cdot \text{diag}(\mathbf{J}_1 \cdot \mathbf{a})^H \\ &\quad \cdot \left(\mathbf{J}_1 \cdot \mathbf{a} (\mathbf{J}_1 \cdot \mathbf{a})^+ - \mathbf{I}_{M-1} \right) \cdot (\mathbf{J}_2 - e^{j\mu} \cdot \mathbf{J}_1) \\ &= \frac{\gamma(M)}{\sqrt{M}} \cdot (\mathbf{J}_1 \cdot \mathbf{a})^+ \cdot (\mathbf{J}_2 / e^{j\mu} - \mathbf{J}_1) - \frac{\mathbf{g}_D^T}{\sqrt{M}} \cdot \text{diag}(\mathbf{a})^H \quad (141) \end{aligned}$$

where the scalar $\gamma(M)$ and the row-vector \mathbf{g}_D^T are defined as

$$\gamma(M) = \mathbf{1}_{1 \times (M-1)} \cdot \mathbf{G}^{-1} \cdot \mathbf{1}_{(M-1) \times 1} \quad (142)$$

$$\mathbf{g}_D^T = \mathbf{1}_{1 \times (M-1)} \cdot \mathbf{G}^{-1} \cdot (\mathbf{J}_2 - \mathbf{J}_1) \in \mathbb{R}^{1 \times M} \quad (143)$$

For $\gamma(M)$ we can show via Lemma 3

$$\gamma(M) = \sum_{m_1=1}^{M-1} \sum_{m_2=1}^{M-1} [\mathbf{G}^{-1}]_{(m_1, m_2)} = \frac{(M-1)M(M+1)}{M^2+11}.$$

Moreover, for the m -th element of the vector \mathbf{g}_D^T , which we denote as $g_{D,m}$ we can show

$$g_{D,m} = \frac{6}{M^2+11} \cdot (2m-M-1), \quad m = 1, 2, \dots, M \quad (144)$$

Collecting our intermediate results, we have shown that $\mathbf{r}_{\text{SLS}}^T$ can be written as

$$\begin{aligned} \mathbf{r}_{\text{SLS}}^T &= \mathbf{r}_{\text{LS}}^T - \Delta \mathbf{r}_{\text{SLS}}^T \\ &= \sqrt{M} \cdot \left(\left(1 - \frac{\gamma(M)}{M}\right) \cdot (\mathbf{J}_1 \mathbf{a})^+ \cdot \left(\frac{\mathbf{J}_2}{e^{j\mu}} - \mathbf{J}_1\right) \right. \\ &\quad \left. + \frac{1}{M} \cdot \mathbf{g}_D^T \cdot \text{diag}(\mathbf{a})^H \right) \end{aligned} \quad (145)$$

where \mathbf{r}_{LS} has been taken from (106). The next step to computing the mean square error is to calculate the squared norm of the vector $\mathbf{W}_{\text{mat}}^T \cdot \mathbf{r}_{\text{SLS}}$. The first few steps in computing this product are very similar to the LS case. Following (109) we find that in the SLS case, the result is again equal to the product of the squared norm the same vector $\tilde{\mathbf{s}}^T$ and the squared norm of a modified vector $\tilde{\mathbf{a}}_{\text{SLS}}^T$, i.e., $\|\mathbf{r}_{\text{SLS}}^T \cdot \mathbf{W}_{\text{mat}}\|_2^2 = \|\tilde{\mathbf{s}}^T\|_2^2 \cdot \|\tilde{\mathbf{a}}_{\text{SLS}}^T\|_2^2$, where

$$\tilde{\mathbf{a}}_{\text{SLS}}^T = \mathbf{r}_{\text{SLS}}^T \cdot \mathbf{Pr}_a^\perp \quad (146)$$

and $\mathbf{r}_{\text{SLS}}^T$ has been computed in (145). Applying similar arguments as in (111), $\tilde{\mathbf{a}}_{\text{SLS}}^T$ can be further simplified into

$$\begin{aligned} \tilde{\mathbf{a}}_{\text{SLS}}^T &= 12 \cdot \frac{\sqrt{M}}{(M-1)(M^2+11)} \cdot [-1, 0, \dots, 0, e^{-(M-1)j\mu}] \\ &\quad + \frac{\sqrt{M}}{M} \cdot \frac{6}{M^2+11} \cdot \\ &\quad [(-M+1), (-M+3)e^{-j\mu}, \dots, (M-1) \cdot e^{-(M-1)j\mu}] \end{aligned}$$

We conclude that the two vectors $\tilde{\mathbf{a}}_{\text{SLS}}^T$ consists of have the same phase in each element and can hence be conveniently combined. When computing the squared norm of $\tilde{\mathbf{a}}_{\text{SLS}}^T$ by summing the squared magnitude of all elements the phase terms cancel which also confirms the intuition the the result should be independent of the particular position μ . We obtain

$$\begin{aligned} \|\tilde{\mathbf{a}}_{\text{SLS}}^T\|_2^2 &= \sqrt{M}^2 \cdot \left(\frac{-12}{(M-1)(M^2+11)} + \frac{6 \cdot (-M+1)}{M(M^2+11)} \right)^2 \\ &\quad + \sum_{m=2}^{M-1} \frac{36}{M(M^2+11)^2} \cdot (-M+2m-1)^2 \\ &\quad + \sqrt{M}^2 \cdot \left(\frac{12}{(M-1)(M^2+11)} + \frac{6 \cdot (M-1)}{M(M^2+11)} \right)^2 \\ &= 12 \cdot \frac{M^4 - 2M^3 + 24M^2 - 22M + 23}{(M^2+11)^2(M-1)^2} \end{aligned} \quad (147)$$

The mean square error is given by (cf. equation (113)) $\mathbb{E}\{(\Delta\mu_{\text{SLS}})^2\} = \frac{\sigma_n^2}{2} \cdot \|\tilde{\mathbf{s}}^T\|_2^2 \cdot \|\tilde{\mathbf{a}}_{\text{SLS}}^T\|_2^2$. Inserting (110) and (147) we have

$$\mathbb{E}\{(\Delta\mu_{\text{SLS}})^2\}$$

$$\begin{aligned} &= \frac{\sigma_n^2}{2} \cdot \frac{12}{MN\hat{P}_T} \cdot \frac{M^4 - 2M^3 + 24M^2 - 22M + 23}{(M^2+11)^2(M-1)^2} \\ &= \frac{\sigma_n^2}{N \cdot \hat{P}_T} \cdot 6 \cdot \frac{M^4 - 2M^3 + 24M^2 - 22M + 23}{M(M^2+11)^2(M-1)^2}, \end{aligned} \quad (148)$$

which is the desired result. \square

APPENDIX H PROOF OF THEOREM 8

A. R-D Standard ESPRIT

The proof for the R-D extension is in fact quite similar to the proof for the 1-D case provided in Section F. In fact, (101) is still valid, the only difference being that $\mathbf{a}(\mu)$ becomes $\mathbf{a}(\mu^{(1)}) \otimes \dots \otimes \mathbf{a}(\mu^{(R)}) = \mathbf{a}(\mu)$. Therefore, the first steps of the derivation can still be performed in the very same way. We obtain the MSE for R-D Standard ESPRIT as

$$\mathbb{E}\left\{\left(\Delta\mu^{(r)}\right)^2\right\} = \frac{\sigma_n^2}{2} \cdot \left\|\mathbf{r}^{(r)T} \cdot \mathbf{W}_{\text{mat}}\right\|_2^2 = \frac{\sigma_n^2}{2} \cdot \|\tilde{\mathbf{s}}\|_2^2 \cdot \|\tilde{\mathbf{a}}^{(r)}\|_2^2, \quad (149)$$

where $\tilde{\mathbf{s}}$ is the same as in the 1-D case (cf. equation (109)) and $\tilde{\mathbf{a}}^{(r)}$ is given by

$$\tilde{\mathbf{a}}^{(r)T} = \sqrt{M} \frac{\mathbf{a}^H \tilde{\mathbf{J}}_1^{(r)H}}{\left\|\tilde{\mathbf{J}}_1^{(r)} \mathbf{a}\right\|_2^2} \left(\frac{\tilde{\mathbf{J}}_2^{(r)}}{e^{j\mu^{(r)}}} - \tilde{\mathbf{J}}_1^{(r)} \right) \cdot \mathbf{Pr}_a^\perp. \quad (150)$$

Since $\tilde{\mathbf{J}}_1^{(r)}$ selects the $M_r - 1$ out of M_r elements in the r -th mode, we have $\left\|\tilde{\mathbf{J}}_1^{(r)} \mathbf{a}\right\|_2^2 = \frac{M}{M_r} \cdot (M_r - 1)$. Moreover, multiplying (150) out and using the fact that \mathbf{a} satisfies the shift invariance equation in the r -th mode, we obtain

$$\begin{aligned} \tilde{\mathbf{a}}^{(r)T} &= \frac{\sqrt{M} \cdot M_r}{M \cdot (M_r - 1)} \cdot \left(\mathbf{a}^H \cdot \tilde{\mathbf{J}}_1^{(r)H} \cdot \tilde{\mathbf{J}}_2^{(r)} / e^{j\mu^{(r)}} \right. \\ &\quad \left. - \mathbf{a}^H \cdot \tilde{\mathbf{J}}_1^{(r)H} \cdot \tilde{\mathbf{J}}_1^{(r)} \right) \end{aligned} \quad (151)$$

Since the the array steering vector \mathbf{a} and the selection matrices $\tilde{\mathbf{J}}_\ell^{(r)}$ can be factored into Kronecker products according to $\mathbf{a} = \mathbf{a}^{(1)} \otimes \dots \otimes \mathbf{a}^{(R)}$ and $\tilde{\mathbf{J}}_\ell^{(r)} = \mathbf{I}_{\prod_{n=1}^{r-1} M_n} \otimes \mathbf{J}_\ell^{(r)} \otimes \mathbf{I}_{\prod_{n=r+1}^R M_n}$, for $\ell = 1, 2$ and $r = 1, 2, \dots, R$, all “unaffected” modes can be factored out of (151) and we have

$$\begin{aligned} \tilde{\mathbf{a}}^{(r)T} &= \frac{\sqrt{M} \cdot M_r}{M \cdot (M_r - 1)} \cdot \left(\mathbf{a}^{(1)} \otimes \dots \otimes \mathbf{a}^{(r-1)} \right)^H \\ &\quad \otimes \left(\tilde{\mathbf{a}}_1^{(r)} - \tilde{\mathbf{a}}_2^{(r)} \right)^T \otimes \left(\mathbf{a}^{(r+1)} \otimes \dots \otimes \mathbf{a}^{(R)} \right)^H \end{aligned} \quad (152)$$

where $\tilde{\mathbf{a}}_1^{(r)}$ and $\tilde{\mathbf{a}}_2^{(r)}$ are given by

$$\begin{aligned} \tilde{\mathbf{a}}_1^{(r)T} &= \mathbf{a}^{(r)H} \cdot \mathbf{J}_1^{(r)H} \cdot \mathbf{J}_2^{(r)} / e^{j\mu^{(r)}} \quad \text{and} \\ \tilde{\mathbf{a}}_2^{(r)T} &= \mathbf{a}^{(r)H} \cdot \mathbf{J}_1^{(r)H} \cdot \mathbf{J}_1^{(r)}. \end{aligned} \quad (153)$$

Following the same reasoning as for (112) we find

$$\tilde{\mathbf{a}}_1^{(r)T} - \tilde{\mathbf{a}}_2^{(r)T} = [-1, 0, \dots, 0, e^{-(M_r-1)j\mu^{(r)}}] \quad (154)$$

Consequently, the desired norm $\|\tilde{\mathbf{a}}^{(r)}\|_2^2$ is directly found to be

$$\begin{aligned}\|\tilde{\mathbf{a}}^{(r)}\|_2^2 &= \frac{M \cdot M_r^2}{M^2 \cdot (M_r - 1)^2} \cdot \left(\prod_{n=1}^{r-1} \|\mathbf{a}^{(n)}\|_2^2 \right) \cdot 2 \cdot \prod_{n=r+1}^R \|\mathbf{a}^{(n)}\|_2^2 \\ &= 2 \cdot \frac{M_r}{(M_r - 1)^2}.\end{aligned}\quad (155)$$

Therefore, the MSE expression for R -D Standard ESPRIT is given by

$$\begin{aligned}\mathbb{E} \left\{ \left(\Delta \mu^{(r)} \right)^2 \right\} &= \frac{\sigma_n^2}{2} \cdot \frac{1}{M \cdot N \cdot \hat{P}_T} \cdot 2 \cdot \frac{M_r}{(M_r - 1)^2} \\ &= \frac{\sigma_n^2}{N \cdot \hat{P}_T} \cdot \frac{M_r}{M \cdot (M_r - 1)^2},\end{aligned}\quad (156)$$

which proves the first part of the theorem. \square

B. R -D Unitary ESPRIT

The second part of the theorem is to prove that in the R -D case the performance of R -D Unitary ESPRIT and R -D Standard ESPRIT are the same as long as a single source is present. However, for this part, no changes have to be made compared to Appendix F-B: As it was shown there, Forward-Backward-Averaging only affects \mathbf{v}_s and has no effect on \mathbf{u}_s or \mathbf{U}_n . Applying the same steps here immediately proves this part of the theorem.

C. Cramér-Rao Bound

The third part is the simplification of the Cramér-Rao Bound. In the R -D case, the CRB is given by

$$\mathbf{C} = \frac{\sigma_n^2}{2 \cdot N} \cdot \text{Re} \left\{ \left[\mathbf{D}^{(R)\text{H}} \cdot \mathbf{Pr}_A^\perp \cdot \mathbf{D}^{(R)} \right] \odot \left(\mathbf{1}_{R \times R} \otimes \hat{\mathbf{R}}_S^T \right) \right\}^{-1}$$

where $\mathbf{Pr}_A^\perp = \mathbf{I}_M - \mathbf{A} \cdot (\mathbf{A}^H \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^H$ and $\mathbf{D}^{(R)} \in \mathbb{C}^{M \times (d \cdot R)}$ contains the partial derivatives of the array steering vectors \mathbf{a}_n with respect to $\mu_n^{(r)}$ for $n = 1, 2, \dots, d$ and $r = 1, 2, \dots, R$. For the special case of a single source, the CRB simplifies into (cf. Appendix F-C)

$$\begin{aligned}\mathbf{C} &= \frac{\sigma_n^2}{2 \cdot N \cdot \hat{P}_T} \cdot \text{Re} \{ \mathbf{J} \}^{-1}, \\ \mathbf{J} &= \mathbf{D}^{(R)\text{H}} \cdot \left(\mathbf{I}_M - \frac{1}{M} \cdot \mathbf{a} \cdot \mathbf{a}^H \right) \cdot \mathbf{D}^{(R)}\end{aligned}\quad (157)$$

The columns of $\mathbf{D}^{(R)} \in \mathbb{C}^{M \times R}$ are given by $\tilde{\mathbf{d}}^{(r)} = \frac{\partial \mathbf{a}}{\partial \mu^{(r)}} \in \mathbb{C}^{M \times 1}$. Using the fact that $\mathbf{a} = \mathbf{a}^{(1)} \otimes \dots \otimes \mathbf{a}^{(R)}$ we obtain

$$\tilde{\mathbf{d}}^{(r)} = \mathbf{a}^{(1)} \otimes \dots \otimes \mathbf{a}^{(r-1)} \otimes \mathbf{d}^{(r)} \otimes \mathbf{a}^{(r+1)} \otimes \dots \otimes \mathbf{a}^{(R)} \quad (158)$$

where $\mathbf{d}^{(r)} = \frac{\partial \mathbf{a}^{(r)}}{\partial \mu^{(r)}} \in \mathbb{C}^{M_r \times 1} = j \cdot [0, e^{j\mu^{(r)}}, 2 \cdot e^{2j\mu^{(r)}}, \dots, (M-1)e^{(M-1)j\mu^{(r)}}]$. Therefore, the elements of the matrix \mathbf{J} are given by

$$[\mathbf{J}]_{(r_1, r_2)} = \tilde{\mathbf{d}}^{(r_1)\text{H}} \cdot \tilde{\mathbf{d}}^{(r_2)} - \frac{1}{M} \cdot \tilde{\mathbf{d}}^{(r_1)\text{H}} \cdot \mathbf{a} \cdot \mathbf{a}^H \cdot \tilde{\mathbf{d}}^{(r_1)} \quad (159)$$

With the help of (158) we find for the diagonal elements ($r_1 = r_2 = r$)

$$\begin{aligned}\tilde{\mathbf{d}}^{(r)\text{H}} \cdot \tilde{\mathbf{d}}^{(r)} &= \frac{M}{M_r} \cdot \left(\sum_{m=0}^{M_r-1} m^2 \right) \\ &= \frac{1}{6} \cdot M \cdot (M_r - 1) \cdot (2M_r - 1)\end{aligned}\quad (160)$$

and similarly

$$\begin{aligned}\tilde{\mathbf{d}}^{(r)\text{H}} \cdot \mathbf{a} &= \frac{M}{M_r} \cdot \left(-j \sum_{m=0}^{M_r-1} m \right) \\ &= -j \cdot M \cdot \frac{1}{2} \cdot (M_r - 1).\end{aligned}\quad (161)$$

Combining these two results we have for $[\mathbf{J}]_{(r, r)}$

$$[\mathbf{J}]_{(r, r)} = \frac{1}{12} \cdot M \cdot (M_r - 1)(M_r + 1). \quad (162)$$

On the other hand, for the off-diagonal elements we obtain

$$\tilde{\mathbf{d}}^{(r_1)\text{H}} \cdot \tilde{\mathbf{d}}^{(r_2)} = \frac{1}{4} M \cdot (M_{r_1} - 1) \cdot M_{r_1} \cdot (M_{r_2} - 1) \cdot M_{r_2}$$

and therefore for $[\mathbf{J}]_{(r_1, r_2)}$, $r_1 \neq r_2$

$$[\mathbf{J}]_{(r_1, r_2)} = \tilde{\mathbf{d}}^{(r_1)\text{H}} \cdot \tilde{\mathbf{d}}^{(r_2)} - \tilde{\mathbf{d}}^{(r_1)\text{H}} \cdot \mathbf{a} \cdot \mathbf{a}^H \cdot \tilde{\mathbf{d}}^{(r_2)} = 0.$$

This shows that \mathbf{J} is diagonal and real-valued. Consequently, the CRB becomes

$$\begin{aligned}\mathbf{C} &= \frac{\sigma_n^2}{2 \cdot N \cdot \hat{P}_T} \cdot \text{Re} \{ \mathbf{J} \}^{-1} = \text{diag} \left(\left[C^{(1)}, \dots, C^{(R)} \right] \right) \\ C^{(r)} &= \frac{\sigma_n^2}{2 \cdot N \cdot \hat{P}_T} \cdot \frac{12}{M \cdot (M_r - 1) \cdot (M_r + 1)} \\ &= \frac{1}{\hat{\rho}} \cdot \frac{6}{M \cdot (M_r - 1) \cdot (M_r + 1)},\end{aligned}\quad (163)$$

which is the desired result. \square

D. R -D Standard Tensor-ESPRIT

The fourth part of the theorem is to show that the MSE of R -D Standard Tensor-ESPRIT is the same as the MSE for R -D Standard ESPRIT for $d = 1$. Since we have only shown the expressions for R -D Standard Tensor-ESPRIT in the special case $R = 2$, we will also assume this case here.

Note that the MSE expression for Tensor-ESPRIT is in fact quite similar to the one for matrix-based ESPRIT with the only difference being that the matrix \mathbf{W}_{mat} is replaced by the matrix \mathbf{W}_{ten} , cf. (35) and (36), respectively.

To simplify this expression for the special case $d = 1$, we express the unfoldings of \mathcal{X}_0 as

$$\begin{aligned}[\mathcal{X}_0]_{(1)} &= \mathbf{a}^{(1)} \cdot \left(\mathbf{a}^{(2)} \otimes \mathbf{s} \right)^T, \quad [\mathcal{X}_0]_{(2)} = \mathbf{a}^{(2)} \cdot \left(\mathbf{s} \otimes \mathbf{a}^{(1)} \right)^T, \\ [\mathcal{X}_0]_{(3)} &= \mathbf{s} \cdot \left(\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)} \right)^T.\end{aligned}$$

Consequently, we can relate the necessary subspaces of the unfoldings of \mathcal{X}_0 to \mathbf{s} and $\mathbf{a}^{(r)}$ via

$$\begin{aligned} \mathbf{u}_1^{[s]} &= \frac{\mathbf{a}^{(1)}}{\sqrt{M_1}}, \quad \mathbf{u}_2^{[s]} = \frac{\mathbf{a}^{(2)}}{\sqrt{M_2}}, \quad \mathbf{u}_3^{[s]} = \frac{\mathbf{s}}{\sqrt{N \cdot \hat{P}_T}} \\ U_1^{[n]} &= \mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp, \quad U_2^{[n]} = \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp \\ \Sigma_1^{[s]} &= \Sigma_2^{[s]} = \Sigma_3^{[s]} = \sqrt{M \cdot N \cdot \hat{P}_T} \\ \mathbf{v}_1^{[s]} &= \frac{(\mathbf{a}^{(2)} \otimes \mathbf{s})^*}{\sqrt{M_2 \cdot N \cdot \hat{P}_T}}, \quad \mathbf{v}_2^{[s]} = \frac{(\mathbf{s} \otimes \mathbf{a}^{(1)})^*}{\sqrt{M_1 \cdot N \cdot \hat{P}_T}}, \\ \mathbf{v}_3^{[s]} &= \mathbf{u}_s = \frac{\mathbf{a}}{\sqrt{M}} \\ V_3^{[n]*} \cdot V_3^{[n]T} &= U_n \cdot U_n^H = \mathbf{Pr}_{\mathbf{a}}^\perp. \end{aligned} \quad (164)$$

Moreover, we have for T_r

$$\begin{aligned} T_r &= \mathbf{u}_r^{[s]} \cdot \mathbf{u}_r^{[s]H} = \mathbf{Pr}_{\mathbf{a}^{(r)}} \quad \text{for } r = 1, 2 \text{ and thus} \\ T_1 \otimes T_2 &= \mathbf{Pr}_{\mathbf{a}^{(1)}} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}} = \mathbf{Pr}_{\mathbf{a}}. \end{aligned} \quad (165)$$

From (164) and (165) it immediately follows that the first term in \mathbf{W}_{ten} cancels as it contains $[T_1 \otimes T_2] \cdot V_3^{[n]*} \cdot V_3^{[n]T}$. We also find $\mathbf{t}_{r,m} = \mathbf{a}^{(r)} \cdot e^{-j\mu^{(r)}(m-1)}/M_r$ for $r = 1, 2$ and $m = 1, 2, \dots, M_r$. To simplify the remaining two terms in \mathbf{W}_{ten} we first simplify some of their components. Using the identity $\mathbf{u}_s = \mathbf{a}/\sqrt{M}$ and the explicit expression for $\mathbf{t}_{r,m}$ it is easy to show that

$$(\mathbf{u}_s^T \otimes \mathbf{I}_M) \cdot \bar{T}_1 = \frac{1}{\sqrt{M}} \left(\mathbf{a}^{(2)T} \otimes \mathbf{a}^{(1)} \otimes \mathbf{I}_{M_2} \right) \quad (166)$$

$$(\mathbf{u}_s^T \otimes \mathbf{I}_M) \cdot \bar{T}_2 = \frac{1}{\sqrt{M}} \left(\mathbf{a}^{(1)T} \otimes \mathbf{I}_{M_1} \otimes \mathbf{a}^{(2)} \right). \quad (167)$$

Moreover, using the relations from (164) we can rewrite $U_r^{[s]*} \Sigma_r^{[s]-1} V_r^{[s]T} \otimes U_r^{[n]} U_r^{[n]H}$ as

$$\begin{aligned} & \left(U_1^{[s]*} \Sigma_1^{[s]-1} V_1^{[s]T} \right) \otimes \left(U_1^{[n]} U_1^{[n]H} \right) \\ &= \frac{1}{MN\hat{P}_T} \cdot \left(\mathbf{a}^{(1)*} \cdot (\mathbf{a}^{(2)} \otimes \mathbf{s})^H \right) \otimes \mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \\ & \left(U_2^{[s]*} \Sigma_2^{[s]-1} V_2^{[s]T} \right) \otimes \left(U_2^{[n]} U_2^{[n]H} \right) \\ &= \frac{1}{MN\hat{P}_T} \cdot \left(\mathbf{a}^{(2)*} \cdot (\mathbf{s} \otimes \mathbf{a}^{(1)})^H \right) \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp. \end{aligned}$$

Combining these intermediate result, the third term in \mathbf{W}_{ten} can be expressed as

$$\begin{aligned} & (\mathbf{u}_s^T \otimes \mathbf{I}_M) \cdot \bar{T}_1 \cdot \left((U_2^{[s]*} \Sigma_2^{[s]-1} V_2^{[s]T}) \otimes (U_2^{[n]} U_2^{[n]H}) \right) \\ &= \frac{M_2}{MN\hat{P}_T\sqrt{M}} \left(\mathbf{s}^H \otimes \mathbf{a}^{(1)H} \otimes \mathbf{a}^{(1)} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp \right) \end{aligned} \quad (168)$$

With similar arguments, the second term in \mathbf{W}_{ten} can be simplified into

$$\begin{aligned} & (\mathbf{u}_s^T \otimes \mathbf{I}_M) \cdot \bar{T}_2 \cdot \left((U_1^{[s]*} \Sigma_1^{[s]-1} V_1^{[s]T}) \otimes (U_1^{[n]} U_1^{[n]H}) \right) \cdot \mathbf{K}_{M_2 \times (M_1 \cdot N)} \\ &= \frac{M_1}{MN\hat{P}_T\sqrt{M}} \left(\mathbf{a}^{(2)H} \otimes \mathbf{s}^H \otimes \mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{a}^{(2)} \right) \cdot \mathbf{K}_{M_2 \times (M_1 \cdot N)} \\ &= \frac{M_1}{MN\hat{P}_T\sqrt{M}} \left(\mathbf{s}^H \otimes \mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{a}^{(2)} \otimes \mathbf{a}^{(2)H} \right). \end{aligned} \quad (169)$$

where the last step is a special case of Property (8) for commutation matrices.

Using (168) and (169) in (36), we obtain

$$\mathbf{W}_{\text{ten}} = \frac{1}{N\hat{P}_T\sqrt{M}} \cdot \mathbf{s}^H \otimes \left(\mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}} + \mathbf{Pr}_{\mathbf{a}^{(1)}} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp \right). \quad (170)$$

Comparing (170) matrix-based counterpart in (150) we find that for a single source, \mathbf{W}_{mat} and \mathbf{W}_{ten} are in fact quite similar, the only difference being that $\mathbf{Pr}_{\mathbf{a}}^\perp$ is replaced by $\mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}} + \mathbf{Pr}_{\mathbf{a}^{(1)}} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp$. Therefore, to show the R -D Standard ESPRIT and R -D Standard Tensor-ESPRIT have the same MSE for $d = 1$, it is sufficient to show that the corresponding terms $\bar{\mathbf{a}}^{(r)}$ are the same, i.e., that

$$\bar{\mathbf{a}}^{(r)} \cdot \mathbf{Pr}_{\mathbf{a}}^\perp = \bar{\mathbf{a}}^{(r)} \cdot \left(\mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}} + \mathbf{Pr}_{\mathbf{a}^{(1)}} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp \right) \quad (171)$$

where $\bar{\mathbf{a}}^{(r)} = \mathbf{a}^H \tilde{\mathbf{J}}_1^{(r)H} \left(\tilde{\mathbf{J}}_2^{(r)} / e^{j\mu^{(r)}} - \tilde{\mathbf{J}}_1^{(r)} \right)$ for $r = 1, 2$. Note that $\bar{\mathbf{a}}^{(r)} \cdot \mathbf{Pr}_{\mathbf{a}}^\perp$ was shown to be equal to (cf. equation (152)) $(\tilde{\mathbf{a}}_1^{(1)} - \tilde{\mathbf{a}}_2^{(1)})^T \otimes \mathbf{a}^{(2)T}$ for $r = 1$ and $\mathbf{a}^{(1)T} \otimes (\tilde{\mathbf{a}}_1^{(2)} - \tilde{\mathbf{a}}_2^{(2)})^T$ for $r = 2$, where $\tilde{\mathbf{a}}_1^{(r)T} = \mathbf{a}^{(r)H} \cdot \mathbf{J}_1^{(r)H} \cdot \mathbf{J}_2^{(r)} / e^{j\mu^{(r)}}$ and $\tilde{\mathbf{a}}_2^{(r)T} = \mathbf{a}^{(r)H} \cdot \mathbf{J}_1^{(r)H} \cdot \mathbf{J}_1^{(r)}$. Expanding the corresponding right-hand side of (171) we have for $r = 1$

$$\begin{aligned} & \bar{\mathbf{a}}^{(r)} \cdot \left(\mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}} + \mathbf{Pr}_{\mathbf{a}^{(1)}} \otimes \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp \right) \\ &= \left(\mathbf{a}^{(1)H} \mathbf{J}_1^{(1)H} \left(\mathbf{J}_2^{(1)} / e^{j\mu^{(1)}} - \mathbf{J}_1^{(1)} \right) \cdot \mathbf{Pr}_{\mathbf{a}^{(1)}}^\perp \right) \otimes \mathbf{a}^{(2)H} \\ &= \left(\tilde{\mathbf{a}}_1^{(1)T} - \tilde{\mathbf{a}}_2^{(1)T} \right) \otimes \mathbf{a}^{(2)H}, \end{aligned}$$

where we have used the fact that $\mathbf{a} = \mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)}$ from its definition, $\mathbf{a}^{(2)H} \cdot \mathbf{Pr}_{\mathbf{a}^{(2)}} = \mathbf{a}^{(2)H}$ and $\mathbf{a}^{(2)H} \cdot \mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp = \mathbf{0}_{1 \times M_2}$ since $\mathbf{Pr}_{\mathbf{a}^{(2)}}$ and $\mathbf{Pr}_{\mathbf{a}^{(2)}}^\perp$ are projectors onto $\mathbf{a}^{(2)}$ and its orthogonal complement, and the identity $\mathbf{J}_2^{(1)} / e^{j\mu^{(1)}} \mathbf{a}^{(1)} - \mathbf{J}_1^{(1)} \mathbf{a}^{(1)} = \mathbf{0}_{(M_2-1) \times 1}$ since $\mathbf{a}^{(1)}$ satisfies the shift invariance equation for $r = 1$. This shows that the left-hand side and the right-hand side of (171) are equal for $r = 1$. The proof for $r = 2$ proceeds in an analogous fashion. Consequently, we have shown that for $d = 1$

$$\mathbf{r}^{(r)T} \cdot \mathbf{W}_{\text{mat}} = \mathbf{r}^{(r)T} \cdot \mathbf{W}_{\text{ten}}, \quad \text{for } r = 1, 2 \quad (172)$$

and hence the MSE for 2-D Standard ESPRIT and 2-D Standard Tensor-ESPRIT are in fact equal. \square

E. R -D Unitary Tensor-ESPRIT

The fifth and final part of the theorem is to show that the MSE for R -D Unitary ESPRIT is again equal to the MSE for R -D Standard ESPRIT in case of a single source. Again, there is no need to derive this in full detail. As it was shown in Appendix F-B, Forward-Backward-Averaging has no effect on \mathbf{u}_s or \mathbf{U}_n but only affects \mathbf{v}_s and \mathbf{V}_n . This carries over to the tensor case where only the quantities involving the symbols are affected. However, since the “symbol part” and the “array part” can always be factorized (cf. equation (170)), the arguments from Appendix F-B can still be applied to prove this part of the theorem. \square

REFERENCES

- [1] T. W. Anderson, "Asymptotic theory for principal component analysis", *Ann. Math. Statist.*, vol. 34, pp. 122–148, 1963.
- [2] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holt, Rhinehart and Winston, New York, 1975.
- [3] L. De Lathauwer, "First-order perturbation analysis of the best rank- (r_1, r_2, r_3) approximation in multilinear algebra", *Journal of Chemometrics*, vol. 18, no. 1, pp. 2–11, 2004.
- [4] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition", *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [5] J. P. Delmas and Y. Meurisse, "On the second-order statistics of the EVD of sample covariance matrices - application to the detection of noncircular or/and NonGaussian components", *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 4017–4023, Aug. 2011.
- [6] A. Eriksson, P. Stoica, and T. Soderstrom, "Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios", *IEE Proceedings-F*, vol. 140, no. 4, Aug. 1993.
- [7] B. Friedlander, "A sensitivity analysis of MUSIC algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, pp. 1740–1751, 1990.
- [8] M. Haardt, "Structured least squares to improve the performance of ESPRIT-type algorithms", *IEEE Transactions on Signal Processing*, vol. 45, pp. 792–799, Mar. 1997.
- [9] M. Haardt and J. A. Nossék, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden", *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1232–1242, May 1995.
- [10] M. Haardt and J. A. Nossék, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems", *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 161–169, Jan. 1998.
- [11] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems", *IEEE Trans. Sig. Proc.*, vol. 56, pp. 3198 – 3213, July 2008.
- [12] M. Haardt, R. S. Thomä, and A. Richter, "Multidimensional high-resolution parameter estimation with applications to channel sounding", in *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Chen, Eds., pp. 255–338. Marcel Dekker, New York, NY, 2004, Chapter 5.
- [13] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise", *IEEE Transactions on Signal Processing*, vol. 39, pp. 892–900, Apr. 1991.
- [14] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 331–341, Apr. 1986.
- [15] H. Krim, P. Forster, and J. G. Proakis, "Operator approach to performance analysis of root-MUSIC and root-min-norm", *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1687–1696, July 1992.
- [16] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, New York Academic Press, 2nd edition, 1978.
- [17] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Unification, simplifications, and observations", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.
- [18] F. Li and R. J. Vaccaro, "Performance degradation of DOA estimators due to unknown noise fields", *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 686–690, Mar. 1992.
- [19] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in singular value decomposition", *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3044–3049, July 2008.
- [20] X. Liu, N. D. Sidiropoulos, and A. Swami, "Blind high-resolution localization and tracking of multiple frequency hopped signals", *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 889–901, Apr. 2002.
- [21] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, John Wiley and Sons, 1995.
- [22] C. P. Mathews, M. Haardt, and M. D. Zoltowski, "Performance analysis of closed-form, ESPRIT based 2-D angle estimator for rectangular arrays", *IEEE Signal Processing Letters*, vol. 3, pp. 124–126, Apr. 1996.
- [23] C. P. Mathews and M. D. Zoltowski, "Performance analysis of the UCA-ESPRIT algorithm for circular ring arrays", *IEEE Transactions on Signal Processing*, vol. 42, no. 9, Sept. 1994.
- [24] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar", *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5693–5705, Nov. 2010.
- [25] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the Total Least Squares ESPRIT algorithm", *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.
- [26] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-Type high resolution estimators for direction finding in correlated and coherent scenes", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. AASP-37, no. 8, Aug. 1989.
- [27] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 532–544, Apr. 1988.
- [28] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. AASP-37, no. 12, Dec. 1989.
- [29] B. D. Rao and K. V. S. Hari, "Performance analysis of Root-MUSIC", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1948, Dec. 1989.
- [30] F. Roemer, H. Becker, M. Haardt, and M. Weis, "Analytical performance evaluation for HOSVD-based parameter estimation schemes", in *Proc. of the IEEE Int. Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP 2009)*, Aruba, Dutch Antilles, Dec. 2009.
- [31] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT - a subspace rotation approach to estimation of parameters of cisoids in noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. AASP-34, pp. 1340–1342, Oct. 1986.
- [32] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 243–258, Mar. 1986.
- [33] S. T. Smith, "Statistical resolution limits and the complexified cramer-rao bound", *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1597–1609, May 2005.
- [34] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.
- [35] P. Stoica and T. Soderstrom, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies", *IEEE Transactions on Signal Processing*, vol. 39, no. 8, Aug. 1991.
- [36] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT", *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 867–881, Apr. 1992.
- [37] A. Thakre, M. Haardt, and K. Giridhar, "Single snapshot r -d unitary esprit using an augmentation of the tensor order", in *Prof. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009)*, Dec. 2009.
- [38] A. Thakre, M. Haardt, and K. Giridhar, "Single snapshot spatial smoothing with improved effective array aperture", *IEEE Signal Processing Letters*, vol. 16, pp. 505–509, June 2009.
- [39] A. Thakre, M. Haardt, F. Roemer, and K. Giridhar, "Tensor-Based spatial smoothing (TB-SS) using multiple snapshots", *IEEE Trans. Sig. Proc.*, Jan. 2010.
- [40] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms", *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2820–2830, Nov. 2002.
- [41] M. D. Zoltowski, G. M. Kautz, and C. P. Mathews, "Performance analysis of eigenstructure based DOA estimators employing conjugate symmetric beamformers", in *IEEE Sixth SP Workshop on Statistical Signal and Array Processing*, Victoria, BC, Canada, Oct. 1992.